

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:30:03 ; Search time 64 Seconds  
(without alignments)  
101.541 Million cell updates/sec

Title: US-10-070-220-13  
Sequence: 1 KGSVWIVGRILLSGRPAIVPRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 581562

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 110   | 100.0       | 23     | 4  | AAB96855 Hepatitis |
| 2          | 99    | 90.0        | 23     | 2  | Aaw40552 Synthetic |
| 3          | 99    | 90.0        | 23     | 4  | Aam52536 Peptide K |
| 4          | 99    | 90.0        | 23     | 4  | Aab67391 KQNS4a pe |
| 5          | 99    | 90.0        | 23     | 4  | Aab66371 Hepatitis |
| 6          | 99    | 90.0        | 23     | 4  | Aag64158 KQNS4a pe |
| 7          | 99    | 90.0        | 23     | 4  | Abg31914 KQNS4a de |
| 8          | 99    | 90.0        | 23     | 5  | Aau76376 Hepatitis |
| 9          | 99    | 90.0        | 23     | 5  | Aae18687 NS4A pept |
| 10         | 99    | 90.0        | 23     | 5  | Abg32508 Peptide K |
| 11         | 99    | 90.0        | 23     | 6  | Abg72264 Hepatitis |
| 12         | 99    | 90.0        | 23     | 7  | Abu61521 Hepatitis |
| 13         | 99    | 90.0        | 23     | 7  | Adc06772 HCV NS4A  |
| 14         | 91    | 82.7        | 28     | 2  | Aaw37386 Hepatitis |
| 15         | 89    | 80.9        | 23     | 2  | Aay15763 Substrate |
| 16         | 89    | 80.9        | 23     | 3  | Aab23810 Synthetic |
| 17         | 89    | 80.9        | 23     | 3  | Aab92236 Virus rel |
| 18         | 78    | 70.9        | 30     | 5  | Aau84710 HCV HepC1 |
| 19         | 74    | 67.3        | 17     | 3  | Aay83775 HCV NS3A  |
| 20         | 74    | 67.3        | 17     | 3  | Aay83773 HCV NS3A  |
| 21         | 74    | 67.3        | 17     | 3  | Aay99552 Hepatitis |
| 22         | 74    | 67.3        | 17     | 3  | Aab97114 Hepatitis |
| 23         | 70    | 63.6        | 16     | 6  | Aao23353 HCV const |
| 24         | 69    | 62.7        | 16     | 3  | Aay54448 Peptide 4 |
| 25         | 69    | 62.7        | 16     | 4  | Aab96851 Hepatitis |

ALIGNMENTS

RESULT 1

AAB96855  
ID AAB96855 standard; peptide; 23 AA.

XX AAB96855;

DT 06-JUL-2001 (first entry)

DE Hepatitis C virus NS2/3 cleavage inhibitory peptide SEQ ID NO: 13.

XX Hepatitis C virus; HCV; NS2; NS3; inhibitory peptide; cleavage;

KW replication inhibition; chimpanzee; human; infection; gene therapy.

XX Hepatitis C virus.

PH Key Location/Qualifiers

FT Modified-site 23

FT /label= OTHER

FT /note= "C-terminal amide"

XX WO200116379-A1.

XX 08-MAR-2001.

XX 25-AUG-2000; 2000WO-US023444.

XX 30-AUG-1999; 99US-0151395P.

XX (MERI ) MERCK & CO INC.

XX Darke PL, Jacobs AR, Kuo LC;

XX WPI, 2001-343059/36.

XX Inhibiting hepatitis C virus (HCV) replication in HCV infected cell, or

XX in a patient or treating a patient for HCV infection comprises inhibiting

XX autocleavage of NS2/3.

XX Claim 29; Page 34; 50pp; English.

XX The present invention describes methods and compositions capable of

XX preventing the replication of hepatitis C virus (HCV), involving

XX administering a compound which inhibits NS2/3 autocleavage. Also provided

XX are peptides capable of inhibiting this cleavage step, of which this

XX sequence is an example. These are useful in the treatment of HCV

XX infection in humans and chimpanzees, and in research applications, for

XX example in studying the stabilisation of NS2/3, the effects of NS2/3 on

XX HCV polyprotein processing and the effects of inhibiting NS2/3

Aao23354 Hepatitis  
Aab05387 NS4a pept  
Aau84709 HCV HepC1  
Aao23358 Hepatitis  
Aao23355 Hepatitis  
Aao23356 Hepatitis  
Aao23357 Hepatitis  
Aaw13792 Hepatitis  
Aab23337 Virus rel  
Aab96853 Hepatitis  
Aab96854 Hepatitis  
Aab96862 Hepatitis  
Aab74387 NS3 prote  
Aaw09243 HCV NS4A  
Aae22013 HCV fragm  
Aay44731 Hepatitis  
Aab96863 Hepatitis  
Aam48242 Hepatitis  
Aab96865 Hepatitis

```

CC autocleavage
XX
SQ Sequence 23 AA;

Query Match          100.0%; Score 110; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILISGRPAIVPRR 23
DB 1 KKGSVVIVGRILISGRPAIVPRR 23

RESULT 2
AAW40552
ID AAW40552 standard; peptide; 23 AA.
XX
AC AAW40552;
XX
DT 20-NOV-1998 (first entry)
XX
DE Synthetic nonstructural peptide SNS4A.
XX
KW Synthetic Hepatitis C nonstructural protein; SNS4A peptide; cofactor;
KW NS3 protease.
XX
OS Hepatitis C virus.
XX
PN WO9811134-A1.
XX
PD 19-MAR-1998.
XX
PF 12-SEP-1997; 97WO-US016182.
XX
PR 12-SEP-1996; 96US-0025274P.
PR 18-OCT-1996; 96US-00731336.
XX
PA (VERT-) VERTEX PHARM INC.
XX
PI Kim JL, Morgenstern KA, Lin C, Fox T, Thomson JA;
XX
DR WPI; 1998-250953/22.
XX
PT New hepatitis C virus crystal compositions - comprising a HCV NS3-like
PT polypeptide complexed with a NS4A-like polypeptide, used particularly for
PT drug design.
XX
PS Claim 4; Page 30; 97pp; English.
XX
CC This is the amino acid sequence of the novel SNS4A (synthetic Hepatitis C
CC nonstructural protein 4A) peptide. It acts as a cofactor for the NS3
CC protease in order to achieve proteolytic processing of Hepatitis C virus
CC (HCV) nonstructural proteins. It is used in the method of the invention
CC as part of a device which can be used to provide information for the
CC design of drugs for the treatment of HCV infection. They can also be used
CC for determining the 3-dimensional structure of molecules or molecular
CC complexes which contain at least some structurally similar features to a
CC HCV NS3 serine protease domain
XX
SQ Sequence 23 AA;

Query Match          90.0%; Score 99; DB 2; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILISGRPAIVPRR 23
DB 1 KKGSVVIVGRILISGRPAIVPRR 23

RESULT 3
AAM52536
ID AAM52536 standard; peptide; 23 AA.
XX
AC AAM52536;
XX
DT 31-JAN-2002 (first entry)
XX
DE Peptide KKS4a used in an enzyme assay.
XX
KW Virucide; pyrrolopyrazinone derivative; Hepatitis C virus inhibitor;
KW nonstructural 3 protease; NS3 protease; viral infection.
XX
OS Synthetic.
XX
PN WO200164678-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006269.
XX
PR 29-FEB-2000; 2000US-0185618P.
XX
PA (DUPO ) DUPONT PHARM CO.
XX
PI Zhang X, Han W;
XX
DR WPI; 2001-656752/75.
XX
PT New pyrrolopyrazinone derivatives useful for treating Hepatitis C virus
PT infection are NS3 protease inhibitors.
XX
PS Disclosure; Page 130; 191pp; English.
XX
CC The present invention relates to a novel pyrrolopyrazinone derivative,
CC its stereoisomer or salt. It was found that the derivative is a Hepatitis
CC C virus (HCV) nonstructural (NS) 3 protease inhibitor. The derivative is
CC useful for the manufacture of a medicament for the treatment of HCV and
CC in therapy for treating HCV infection. The present peptide was used to
CC illustrate the present invention
XX
SQ Sequence 23 AA;

Query Match          90.0%; Score 99; DB 4; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILISGRPAIVPRR 23
DB 1 KKGSVVIVGRILISGRPAIVPRR 23

RESULT 4
AAB67391
ID AAB67391 standard; peptide; 23 AA.
XX
AC AAB67391;
XX
DT 26-APR-2001 (first entry)
XX
DE KKS4a peptide.
XX
KW Lactam; hepatitis C virus; HCV; NS3 protease.
XX
OS Synthetic.
XX
PN WO200107407-A1.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US020189.
XX
PR 26-JUL-1999; 99US-0145631P.
XX
PA (DUPO ) DU PONT PHARM CO.

```

XX Priestley ES, Decicco CP;  
XX WPI; 2001-159696/16.  
XX  
XX New lactam derivatives are hepatitis C virus NS3 protease inhibitors  
XX useful for treating HCV infections.  
XX  
XX Example 26; Page 100; 130pp; English.  
XX  
XX The present invention relates to Lactam derivatives. These derivatives  
XX may be used for treating hepatitis C virus (HCV) infection. They can also  
XX be used for inhibiting HCV in a body fluid sample and as a standard or  
XX reagent in a test or assay for determining the ability of a potential  
XX pharmaceutical to inhibit HCV NS3 protease and/or HCV growth  
XX  
XX Sequence 23 AA;  
SQ  
Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KGSVVIVGRILSGRPAIVPRR 23  
DB 1 KGSVVIVGRILSGRPAIVPRR 23  
RESULT 5  
ID AAB66371 standard; peptide; 23 AA.  
XX AAB66371;  
XX  
XX 09-APR-2001 (first entry)  
XX  
XX Hepatitis C virus protease inhibitor related peptide #1.  
XX  
XX Hepatitis C virus; protease; boronic acid; inhibitor; liver cirrhosis;  
XX liver cancer; NS3; antiviral agent.  
XX  
XX Unidentified.  
XX  
XX WO200102424-A2.  
XX  
XX 11-JAN-2001.  
XX  
XX 07-JUL-2000; 2000WO-US018655.  
XX  
XX 07-JUL-1999; 99US-0142561P.  
XX  
XX (DUPO ) DU PONT PHARM CO.  
XX  
XX Kettner CA, Jagannathan S, Forsyth TP;  
XX  
XX WPI; 2001-103001/11.  
XX  
XX New boronic acid derivatives, optionally containing peptides, used to  
XX treat hepatitis C infections, are hepatitis C viral protease inhibitors.  
XX  
XX Example 60; Page 208; 258pp; English.  
XX  
XX The present invention provides a number of boronic acid derivatives which  
XX act as inhibitors of the hepatitis C virus NS3 protease enzyme. They can  
XX be used to treat infection by the virus, which can cause liver cirrhosis  
XX and liver cancer  
XX  
XX Sequence 23 AA;  
SQ  
Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KGSVVIVGRILSGRPAIVPRR 23

DB 1 KGSVVIVGRILSGRPAIVPRR 23  
RESULT 6  
ID AAG64158 standard; peptide; 23 AA.  
XX AAG64158;  
XX  
XX 19-OCT-2001 (first entry)  
XX  
XX KNS4a peptide.  
XX  
XX Hepatitis C virus; HCV; NS3 protease; alpha-ketoamide inhibitor;  
XX virucide; hepatotropic; antiinflammatory; viral infection; KNS4a.  
XX  
XX Synthetic.  
XX  
XX WO200140262-A1.  
XX  
XX 07-JUN-2001.  
XX  
XX 01-DEC-2000; 2000WO-US032677.  
XX  
XX 03-DEC-1999; 99US-0168998P.  
XX  
XX (DUPO ) DU PONT PHARM CO.  
XX  
XX Han W;  
XX  
XX WPI; 2001-464936/50.  
XX  
XX New ketoamide derivatives useful for treating infections e.g. hepatitis C  
XX virus.  
XX  
XX Disclosure; Page 195; 282pp; English.  
XX  
XX The invention relates to novel ketoamide and ketoester derivatives for  
XX use as inhibitors of hepatitis C virus (HCV) NS3 protease inhibitors. The  
XX compounds are useful for treating viral infections e.g. hepatitis C  
XX virus. The present sequence was used in an experiment measuring the  
XX effect of an inhibitor on the rate of hydrolysis of an ester substrate  
XX  
XX Sequence 23 AA;  
SQ  
Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KGSVVIVGRILSGRPAIVPRR 23  
DB 1 KGSVVIVGRILSGRPAIVPRR 23  
RESULT 7  
ID ABG31914 standard; peptide; 23 AA.  
XX ABG31914;  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX KNS4a peptide.  
XX  
XX HCV; hepatitis C; imidazolidinone; serine protease inhibitor; virucide;  
XX hepatotropic; antiinflammatory; NS3 protease; KNS4a; growth inhibitor;  
XX viral infection; blood plasma processing.  
XX  
XX Synthetic.  
XX  
XX WO200248157-A2.  
XX

PD 20-JUN-2002.  
 XX 12-DEC-2001; 2001WO-US047916.  
 PF 13-DEC-2000; 2000US-0255168P.  
 XX (BRIM ) BRISTOL MYERS SQUIBB PHARMA CO.  
 XX Han Q;  
 PI WPI; 2002-599498/64.  
 DR New imidazolidinones useful as serine protease inhibitors in the  
 PT treatment of e.g. viral infection.  
 XX Example 20; Page 112; 173pp; English.  
 XX This invention relates to novel imidazolidinones or their stereoisomers,  
 CC salts or prodrugs which are useful as serine protease inhibitors. The  
 CC imidazolidinones of the invention may have virucide, hepatotropic, or  
 CC anti-inflammatory activities and may be used as a serine protease  
 CC inhibitor (preferably Hepatitis C virus (HCV) NS3 protease inhibitor) or  
 CC a HCV growth inhibitor. Compounds of the invention are useful for  
 CC treating viral infection e.g. Hepatitis C virus (HCV) infection and as a  
 CC reagent used as inhibitors of HCV protease in the processing of blood  
 CC plasma for diagnostic and other commercial purposes. The imidazolidinones  
 CC of the invention inhibit HCV NS3 protease and/or HCV growth and thus can  
 CC be used in the blood plasma assay. The present sequence represents the  
 CC KEN4a peptide used in enzyme assay experiments in the examples of the  
 CC specification  
 XX SQ Sequence 23 AA;  
 Query Match 90.0%; Score 99; DB 5; Length 23;  
 Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 DB 1 KKGSVVIVGRILSGRPAIVPRR 23  
 RESULT 8  
 AAU76376  
 ID AAU76376 standard; peptide; 23 AA.  
 XX AAU76376;  
 AC AAU76376;  
 DT 08-MAY-2002 (first entry)  
 XX Hepatitis C virus (non-structural protein) NS4A peptide sequence.  
 DE Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
 XX immunosay solid support; multiple epitope fusion antigen; MEFA;  
 KW non-structural protein; NS4A.  
 KW Hepatitis C virus.  
 OS WO200196870-A2.  
 XX 20-DEC-2001.  
 XX 14-JUN-2001; 2001WO-US019156.  
 XX 15-JUN-2000; 2000US-0212082P.  
 XX 02-APR-2001; 2001US-0280811P.  
 XX 02-APR-2001; 2001US-0280867P.  
 XX (CHIR ) CHIRON CORP.  
 XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
 PI Medina-Selby A;  
 XX WPI; 2002-179522/23.

DR WPI; 2002-090228/12.  
 XX Immunoassay solid support, useful for detecting hepatitis C virus  
 PT infection in biological sample, comprises HCV NS3/4a conformational  
 PT epitope and multiple epitope fusion antigen bound to the support.  
 XX Example 3; Page 48; 92pp; English.  
 XX The present invention relates to a new immunoassay solid support  
 CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a  
 CC conformational epitope and a multiple epitope fusion antigen (MEFA),  
 CC bound to the support. The NS3/4a conformational epitope and/or MEFA  
 CC reacts specifically with anti-HCV antibodies present in a biological  
 CC sample from an HCV-infected individual. The immunoassay of the invention  
 CC is useful for detecting hepatitis C virus infection in a biological  
 CC sample. The method of the invention provides a sensitive, accurate  
 CC diagnostic and prognostic tool to provide adequate patient care and to  
 CC prevent transmission of HCV by blood and by blood products, or by  
 CC personal contact. Use of NS3/4a conformational epitope in combination  
 CC with MEFA, provides a sensitive and reliable method for detecting early  
 CC HCV seroconversion. Use of MEFA has the added advantages of decreasing  
 CC masking problems, improving sensitivity in detecting antibodies by  
 CC allowing a greater number of epitopes on a unit surface area of  
 CC substrate, and improving substrate. Detection accuracy is increased and  
 CC the incidence of false results is reduced because of the identification  
 CC and the use of highly immunogenic HCV antigens which are present during  
 CC the early stages of HCV seroconversion. The present amino acid sequence  
 CC represents the non-structural protein NS4A peptide sequence. The peptide  
 CC was used in the invention for the purification of NS3/4a conformational  
 CC epitope  
 XX SQ Sequence 23 AA;  
 Query Match 90.0%; Score 99; DB 5; Length 23;  
 Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 DB 1 KKGSVVIVGRILSGRPAIVPRR 23  
 RESULT 9  
 AAU18687  
 ID AAU18687 standard; peptide; 23 AA.  
 XX AAU18687;  
 AC AAU18687;  
 DT 17-MAY-2002 (first entry)  
 XX NS4A peptide used to purify NS3/4a conformational epitope.  
 DE Hepatitis C virus; NS3/4a antibody; HCV infection; NS4A peptide.  
 XX Unidentified.  
 OS WO200196875-A2.  
 XX 20-DEC-2001.  
 XX 14-JUN-2001; 2001WO-US019369.  
 XX 15-JUN-2000; 2000US-0212082P.  
 XX 02-APR-2001; 2001US-0280811P.  
 XX 02-APR-2001; 2001US-0280867P.  
 XX (CHIR ) CHIRON CORP.  
 XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
 PI Medina-Selby A;  
 XX WPI; 2002-179522/23.

PT Immunoassay solid support useful for detecting hepatitis C virus  
 PT infection in a biological sample, comprises at least one of HCV anti-core  
 PT antibody and HCV NS3/4a epitope, bound to the support.  
 XX  
 XX  
 PS Example 2; Page 50; 87pp; English.

CC The present invention relates to hepatitis C virus (HCV) core antigen and  
 CC NS (nonstructural) 3/4a antibody combination assay that can detect both  
 CC HCV antigens and antibodies present in a sample using a single solid  
 CC matrix as well as immunoassay solid supports for use in the assay. The  
 CC solid support is useful for detecting HCV infection in a biological  
 CC sample. The present sequence is NS4a peptide which is used to purify  
 CC NS3/4a conformational epitope in the exemplification of the invention  
 XX  
 XX

SQ Sequence 23 AA;

Query Match 90.0%; Score 99; DB 5; Length 23;

Best Local Similarity 78.3%; Pred. No. 1.6e-08; Length 23;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 |||||:||||:||||:||||:||||:  
 Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 10

ABG32508  
 ID ABG32508 standard; peptide; 23 AA.

AC ABG32508;

DT 15-NOV-2002 (first entry)

DE Peptide KNNS4a for HCV NS3 protease kinetic assay.

XX NS3; HCV; protease; HCV infection; hepatitis; cirrhosis; liver cancer;  
 KW pyrimidinone; serine protease inhibitor; virucide; hepatotropic;  
 KW antiinflammatory; blood plasma processing; KNNS4a.  
 XX Synthetic.

OS

XX WO200248116-A2.

FN 20-JUN-2002.

PD 12-DEC-2001; 2001WO-US047911.

PF 13-DEC-2000; 2000US-0255290P.

XX (BRIM ) BRISTOL-MYERS SQUIBB PHARMA CO.

PA Glunz PW, Doubt BD, Han W;

PI WPI; 2002-627251/67.

XX New pyrimidinones useful as serine protease inhibitors in the treatment  
 of e.g. viral infection.

PS Example 140; Page 192; 270pp; English.

XX The invention relates to pyrimidinones of a formula given in the claims  
 of the specification, their stereoisomers, salts and prodrugs. In assays,  
 CC the pyrimidinone compounds inhibited Hepatitis C virus (HCV) NS3 protease  
 CC with IC<sub>50</sub> values of less than 100 micro M. The compounds are useful for  
 CC treating viral infection e.g. HCV infection (the causative agent of acute  
 CC hepatitis and associated with cirrhosis and liver cancer) and as a  
 CC reagent used as inhibitors of HCV protease in the processing of blood  
 CC plasma for diagnostic and other commercial purposes. The present sequence  
 CC is a peptide, KNNS4a, used in an NS3 kinetic assay  
 XX

SQ Sequence 23 AA;

Query Match 90.0%; Score 99; DB 5; Length 23;

Best Local Similarity 78.3%; Pred. No. 1.6e-08;

Best Local Similarity 78.3%; Pred. No. 1.6e-08;

Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 |||||:||||:||||:||||:||||:  
 Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 11

ABG72264

ID ABG72264 standard; peptide; 23 AA.

XX AC ABG72264;

DT 06-MAR-2003 (first entry)

DE Hepatitis C Virus type-1 (HCV-1) NS4a peptide.

XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;  
 KW NS3/4a conformational epitope; multiple epitope fusion antigen; MEFA;  
 KW anti-HCV antibody; NS3/4a conformational antigen; HCV infection;  
 KW E2 hypervariable region.

XX Hepatitis C virus type 1.

OS US2002146685-A1.

FN 10-OCT-2002.

PD 14-JUN-2001; 2001US-00881654.

PF 15-JUN-2000; 2000US-0212082P.

PR 02-APR-2001; 2001US-0280811P.

PR 02-APR-2001; 2001US-0280867P.

XX (CHIE//) CHIEN D Y.

PA (ARCA//) ARCANDEL P.

PA (TAND//) TANDESKE L.

PA (GEOR//) GEORGE-NASCIMENTO C.

PA (COIT//) COIT D.

PA (MEDI//) MEDINA-SELBY A.

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

PI Medina-Selby A;

PI WPI; 2003-147573/14.

DR Immunoassay solid support for detecting Hepatitis C Virus infection in

PT biological samples, comprises Hepatitis C Virus conformational epitope

PT and multiple epitope fusion antigen.

XX Example 3; Page 17; 45pp; English.

PS The present invention relates to immunoassays comprising Hepatitis C  
 XX Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
 CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
 CC multiple epitope fusion antigen react with anti-HCV antibodies present in  
 CC a biological sample from an HCV-infected individual. The immunoassays and  
 CC methods of the invention are useful for detecting HCV infection in a  
 CC biological sample. The inventive immunoassay solid support provides a  
 CC sensitive and reliable method for detecting early HCV seroconversion. The  
 CC assays can detect HCV infection caused by any six known genotypes of HCV.  
 CC The use of the multiple epitope fusion proteins decreases masking  
 CC problems, improves sensitivity in detecting antibodies by allowing a  
 CC greater number of epitopes on a unit area of substrate, and improves  
 CC selectivity. The present sequence representing HCV type 1 (HCV-1) NS4a  
 CC peptide is used in a protease enzyme activity assay in the examples of  
 CC the present invention  
 XX

SQ Sequence 23 AA;

Query Match 90.0%; Score 99; DB 6; Length 23;

Best Local Similarity 78.3%; Pred. No. 1.6e-08;

Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVWIVGRILSGRPAIVPRR 23  
 PD |||||:||||:||||:||||:||||:  
 Db 1 KKGSVWIVGRIVLSGKPAIIPKK 23

## RESULT 12

ABU61521  
 ID ABU61521 standard; peptide; 23 AA.

XX AC ABU61521;

XX DT 08-AUG-2003 (first entry)

XX DE Hepatitis C virus NS3 protease enzyme assay associated peptide #2.

XX KW Hepatitis C virus; NS3 protease; hepatotropic; virucide;

XX KW antiinflammatory; serine protease inhibitor; hepatitis C virus;

XX KW NS3 Protease Inhibitor; viral infection; elastase; chymotrypsin.

XX OS Synthetic.

XX PN US2002177725-A1.

XX PD 28-NOV-2002.

XX PF 28-OCT-2001; 2001US-00039317.

XX PR 23-OCT-2000; 2000US-0242557P.

XX PA (PRIE/) PRIESTLEY E S.

XX PI Priestley ES;

XX DR WPI; 2003-465950/44.

XX PT New class of peptides are inhibitors of hepatitis C virus NS3 protein,  
 PT useful for treating e.g. hepatitis.

XX PS Example 39; Page 39; 54pp; English.

XX CC The invention describes novel peptide useful for treating viral  
 CC infections e.g. Hepatitis C virus (HCV) infections. The compounds have  
 CC HCV NS3 protease inhibitory selectivity over inhibition of elastase and  
 CC chymotrypsin. This is the amino acid sequence of a modified peptide  
 CC associated with an assay of NS3 protease activity

XX SQ Sequence 23 AA;

Query Match 90.0%; Score 99; DB 7; Length 23;  
 Best Local Similarity 78.3%; Pred. No. 1.6e-08;

Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVWIVGRILSGRPAIVPRR 23

Db 1 KKGSVWIVGRIVLSGKPAIIPKK 23

## RESULT 13

ADC06772  
 ID ADC06772 standard; peptide; 23 AA.

XX AC ADC06772;

XX DT 18-DEC-2003 (first entry)

XX DE HCV NS4A peptide used to monitor protease enzyme activity.

XX KW immunoassay solid support; HCV; NS3/4a; non-structural;

XX KW non-A, non-B hepatitis; NANB; NS4A; NS3 cofactor.

XX OS Hepatitis C virus.

XX PN US2002192639-A1.

XX PD 19-DEC-2002.

XX PF 14-JUN-2001; 2001US-00881239.

XX PR 15-JUN-2000; 2000US-0212082P.

XX PR 02-APR-2001; 2001US-0280811P.

XX PR 02-APR-2001; 2001US-0280867P.

XX PA (CHIE/) CHIEN D Y.

XX PA (ARCA/) ARCANGEL P.

XX PA (TAND/) TANDESKE L.

XX PA (GEOR/) GEORGE-NASCIMENTO C.

XX PA (COIT/) COIT D.

XX PA (MEDI/) MEDINA-SELBY A.

XX PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

XX PI Medina-Selby A;

XX DR WPI; 2003-644609/61.

XX PT Immunoassay solid support for detecting hepatitis C virus infection in  
 PT biological samples, comprises a hepatitis C virus anti-core antibody and  
 PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
 PT antibody.

XX PS Example 3; Page 18; 40pp; English.

XX CC The invention relates to a novel immunoassay solid support comprising at  
 CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
 CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.  
 CC The system of the invention may be useful for detecting HCV infection in  
 CC a biological sample and for treating or detecting non-A, non-B hepatitis  
 CC (NANB hepatitis). The current sequence is that of the HCV NS4A (NS3  
 CC cofactor) peptide of the invention which was used to monitor protease  
 CC enzyme activity.

XX SQ Sequence 23 AA;

Query Match 90.0%; Score 99; DB 7; Length 23;

Best Local Similarity 78.3%; Pred. No. 1.6e-08;

Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVWIVGRILSGRPAIVPRR 23

Db 1 KKGSVWIVGRIVLSGKPAIIPKK 23

## RESULT 14

AAW37386  
 ID AAW37386 standard; peptide; 28 AA.

XX AC AAW37386;

XX DT 27-AUG-2003 (revised)

XX DT 11-MAR-1998 (first entry)

XX DE Hepatitis C virus NS4-I2 protein 1678-1705.

XX KW Hepatitis C virus; HCV; chimeric; antigen; detection; core region;

XX KW epitope; NS3; NS4; infection.

XX OS Hepatitis C virus.

XX PN JP09278794-A.

XX PD 28-OCT-1997.

XX PF 10-FEB-1997; 97JP-00027015.

XX PR 09-FEB-1996; 96JP-00024045.

XX (TOFU ) TONEN CORP.  
XX PA  
XX DR WPI; 1998-022248/03.  
XX  
XX PT New chimaeric peptide antigen derived from hepatitis C virus protein -  
XX useful for detecting HCV infections.  
XX  
XX PS Disclosure; Page 26; 30pp; Japanese.  
XX  
XX CC The present sequence represents a Hepatitis C virus (HCV) protein  
CC sequence from the disclosure of the present specification. The present  
CC specification describes a chimeric HCV peptide antigen which comprises at  
CC least 2 peptide epitope regions from the HCV polypeptide core region, 2  
CC peptide epitope regions from the NS3 region and at least 2 peptide  
CC epitope regions from the NS4 region. The antigen binds specifically with  
CC an antibody produced by a human infected by HCV. The peptide can detect a  
CC wide range of HCV infections with high sensitivity. (Updated on 27-AUG-  
CC 2003 to correct OS field.)  
XX  
XX SQ Sequence 28 AA;  
XX  
XX Query Match 82.7%; Score 91; DB 2; Length 28;  
XX Best Local Similarity 85.7%; Pred. No. 3.6e-07;  
XX Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 3 GSVVIVGRILLSGRPAIVPRR 23  
XX ||||| ||||| ||||| :||  
XX DB 1 GSVVIVGRILLSGRPAIVPDR 21  
XX  
XX RESULT 15  
XX AAY15763  
XX ID AAY15763 standard; peptide; 23 AA.  
XX  
XX AC AAY15763;  
XX  
XX DT 28-JUL-1999 (first entry)  
XX  
XX DE Substrate peptide for HCV protease.  
XX  
XX KW Substrate peptide; MBP-NS5A/5B-ADK; enzyme assay; hepatitis C virus; HCV;  
XX serine protease; maltose bound protein; Escherichia coli.  
XX  
XX OS Hepatitis C virus.  
XX  
XX FN JP11124400-A.  
XX  
XX PD 11-MAY-1999.  
XX  
XX PF 17-OCT-1997; 97JP-00285700.  
XX  
XX PR 17-OCT-1997; 97JP-00285700.  
XX  
XX PA (NIHA ) JAPAN ENERGY CORP.  
XX  
XX DE WPI; 1999-341641/29.  
XX  
XX PT New substrate peptide - useful for enzymic assay of hepatitis C virus  
XX derived serine protease.  
XX  
XX PS Disclosure; Page 2; 12pp; Japanese.  
XX  
XX CC The specification describes a substrate peptide, designated MBP-NS5A/5B-  
XX ADK, that is used in enzyme assays of hepatitis C virus (HCV) derived  
XX serine protease. The substrate peptide comprises a peptide chain bound to  
XX the C-terminal of maltose bound protein derived from Escherichia coli,  
XX and labelled on one of five Lys residues at the C-terminal, especially  
XX with 14C labelled acetyl group on the epsilon amino group of the Lys or  
XX with a fluorescence label of fluorescein isothiocyanate (FITC). The  
XX present sequence represents a substrate peptide for HCV protease  
XX  
XX SQ Sequence 23 AA;  
XX

Query Match 80.9%; Score 89; DB 2; Length 23;  
Best Local Similarity 94.7%; Pred. No. 6e-07;  
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSVVIVGRILLSGRPAIVP 21  
||| ||| ||| ||| ||| :||  
DB 4 GSVVIVGRILLSGRPAVVP 22

Search completed: July 6, 2004, 09:42:05  
Job time : 66 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:35:43 ; Search time 21 Seconds  
(without alignments)  
105.353 Million cell updates/sec

Title: US-10-070-220-13  
Perfect score: 110  
Sequence: 1 KKGSVIVGRILLSGRPAIVPRR 23  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 6282

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 36    | 32.7        | 19     | 2     | A84022      |
| 2          | 35    | 31.8        | 16     | 4     | A33171      |
| 3          | 30    | 27.3        | 21     | 2     | P00145      |
| 4          | 30    | 27.3        | 29     | 4     | I58970      |
| 5          | 29    | 26.4        | 25     | 2     | H4710       |
| 6          | 28    | 25.5        | 22     | 2     | S68900      |
| 7          | 28    | 25.5        | 26     | 2     | I45087      |
| 8          | 27.5  | 25.0        | 24     | 2     | E45087      |
| 9          | 27.5  | 25.0        | 26     | 1     | MEHBCI      |
| 10         | 27.5  | 25.0        | 27     | 1     | MEHBCI      |
| 11         | 27.5  | 25.0        | 30     | 2     | H57647      |
| 12         | 27    | 24.5        | 16     | 2     | A60551      |
| 13         | 27    | 24.5        | 20     | 2     | A56899      |
| 14         | 27    | 24.5        | 20     | 2     | S43635      |
| 15         | 27    | 24.5        | 23     | 2     | S11811      |
| 16         | 27    | 24.5        | 23     | 2     | A19828      |
| 17         | 27    | 24.5        | 24     | 2     | S53793      |
| 18         | 26    | 23.6        | 26     | 1     | MEHBCD      |
| 19         | 26    | 23.6        | 27     | 2     | A95070      |
| 20         | 26    | 23.6        | 28     | 1     | LEPBUT      |
| 21         | 26    | 23.6        | 28     | 2     | AG0516      |
| 22         | 25.5  | 23.2        | 26     | 2     | S59906      |
| 23         | 25    | 22.7        | 13     | 2     | A39836      |
| 24         | 25    | 22.7        | 15     | 2     | S4712       |
| 25         | 25    | 22.7        | 20     | 2     | S65399      |
| 26         | 25    | 22.7        | 20     | 2     | B43534      |
| 27         | 25    | 22.7        | 24     | 2     | B42224      |
| 28         | 25    | 22.7        | 28     | 2     | A61529      |
| 29         | 25    | 22.7        | 28     | 2     | A61322      |

30 25 22.7 28 2 PH1335 Ig heavy chain DJ  
31 25 22.7 29 2 I50382 c-mil protein - ch  
32 25 22.7 30 2 S05223 photosystem I 6.5K  
33 24 21.8 17 2 B20242 pyruvate kinase (E  
34 24 21.8 17 2 A34835 ribosomal protein  
35 24 21.8 18 2 B57789 kidney stone matri  
36 24 21.8 20 2 S17501 glutaminase - Alca  
37 24 21.8 20 2 JP0056 ribosomal protein  
38 24 21.8 25 2 D41575 bombinin-like pept  
39 24 21.8 25 2 A60921 uroepithelial cell  
40 24 21.8 25 2 S03456 T-cell receptor al  
41 24 21.8 26 2 A39769 N-acetylglucosamin  
42 24 21.8 27 2 PC4234 hypothetical prote  
43 24 21.8 28 2 A48855 calcium channel al  
44 24 21.8 29 2 S08088 gene VII protein -  
45 24 21.8 30 2 S74112 proline-rich antib

## ALIGNMENTS

## RESULT 1

A84022  
hypothetical protein BH2977 [imported] - Bacillus halodurans (strain C-125)  
C/Species: Bacillus halodurans  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C/Accession: A84022  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: A84022  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-19 <STO>  
A/Cross-references: GB:AP001517; GB:BA000004; NID:gl0175500; PIDN:BA060696.1; GSPDB:GNOC  
A/Experimental source: strain C-125  
C/Genetics:  
A/Gene: BH2977

Query Match 32.7%; Score 36; DB 2; Length 19;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKGSVVIVG 9  
|||:|  
DB 6 KKGSIIFVG 14

## RESULT 2

A33171  
hypothetical protein lpxA 3'-region - Escherichia coli (fragment)  
C/Species: Escherichia coli  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C/Accession: A33171  
R/Coleman, J.; Raetz, C.R.H.  
J. Bacteriol. 170, 1268-1274, 1988  
A/Title: First committed step of lipid A biosynthesis in Escherichia coli: sequence of t  
A/Reference number: A33171; MUID:88139188; PMID:3277952  
A/Accession: A33171  
A/Molecule type: DNA  
A/Residues: 1-16 <COL>  
A/Cross-references: GB:M19334; GB:M18265; GB:M18266; NID:G450760; DBUJ:D89336; NID:gl208  
C/Comment: This is the hypothetical translation of a sequence that was not reported as a

Query Match 31.8%; Score 35; DB 4; Length 16;  
Best Local Similarity 57.1%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 GSVIVGRILLSGR 16  
|||:|  
DB 1 GRVIVWSVILGR 14

```

RESULT 3
PQ0145
Glucan endo-1,3-beta-D-glucosidase (EC 3.2.1.39) 2 - common tobacco (fragment)
N:Alternate names: endo-1,3-glucanase; laminarinase
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
C:Accession: PQ0145
R:Neale, A.D.; Wahleithner, J.A.; Lund, M.; Bonnett, H.T.; Kelly, A.; Meeks-Wagner, D.R.
Plant Cell 2, 673-684, 1990
A:Title: Chitinase, beta-1,3-glucanase, osmotin, and extensin are expressed in tobacco
A:Reference number: JQ0993; MUID:92404740; PMID:2152343
A:Accession: PQ0145
A:Molecule type: mRNA
A:Residues: 1-21 <NEA>
A:Cross-references: GB:S44871; NID:9256138; PIDN:AAB23377.1; PID:9256139
A:Experimental source: thin cell layer, cv. Samsun NN
C:Comment: This protein is expressed mainly in roots.
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match      27.3%; Score 30; DB 2; Length 21;
Best Local Similarity 71.4%; Pred. No. 4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 17 PAIVPRR 23
DB 13 PSIVPKR 19

RESULT 4
I58970
Hypothetical protein NAD4 locus - rat mitochondrion (fragment)
C:Species: Mitochondrion Rattus norvegicus (Norway rat)
C:Date: 30-Oct-1998 #sequence_revision 30-Oct-1998 #text_change 20-Apr-2000
C:Accession: I58970
R:Castora, F.J.; Arnheim, N.; Simpson, M.V.
Proc. Natl. Acad. Sci. U.S.A. 77, 6415-6419, 1980
A:Title: Mitochondrial DNA polymorphism: evidence that variants detected by restriction
A:Reference number: I58970; MUID:81101062; PMID:6256736
A:Accession: I58970
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-29 <CAS>
A:Cross-references: GB:M10477; NID:9343183; PIDN:AAB59710.1; PID:9829031
C:Comment: The nucleotide sequence is from the NAD4 gene. This translation is in a reading
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Keywords: mitochondrion

Query Match      27.3%; Score 30; DB 4; Length 29;
Best Local Similarity 44.4%; Pred. No. 5.5e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 5 VVIVGRILSGRPAIVPR 22
DB 7 VAINGKIICNPAMIVR 24

RESULT 5
H64710
Hypothetical protein HPI528 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: H64710
R:Tomb, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khairak, H.G.; Glodek, A.; McKen-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: H64710

```

```

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-25 <TCM>
A:Cross-references: GB:AE000651; GB:AE000511; NID:92314708; PIDN:AAD08577.1; PID:9231471
A:Query Match      26.4%; Score 29; DB 2; Length 25;
Best Local Similarity 45.5%; Pred. No. 6.7e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KKGSVVIVGR 11
DB 15 KKGSLVIVNRL 25

RESULT 6
S68900
xanthine dehydrogenase (EC 1.1.1.204) 82.4K chain - Veillonella atypica (fragment)
C:Species: Veillonella atypica
C:Date: 23-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997
C:Accession: S68900
R:Gremer, L.; Meyer, O.
Eur. J. Biochem. 238, 862-866, 1996
A:Title: Characterization of xanthine dehydrogenase from the anaerobic bacterium Veillon
A:Reference number: S68900; MUID:96300255; PMID:8706691
A:Accession: S68900
A:Molecule type: protein
A:Residues: 1-22 <GRE>
A:Experimental source: DSM 1399
C:Keywords: 2Fe-2S; FAD; flavoprotein; heterotrimer; iron-sulfur protein; metalloprotein

Query Match      25.5%; Score 28; DB 2; Length 22;
Best Local Similarity 71.4%; Pred. No. 8.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 12 ILSGRPA 18
DB 16 ILSGKPS 22

RESULT 7
I45087
cysteine proteinase homolog (clone PCR33) - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: I45087
R:Petanceska, S.; Devi, L.
J. Biol. Chem. 267, 26038-26043, 1992
A:Title: Sequence analysis, tissue distribution, and expression of rat cathepsin S.
A:Reference number: A45087; MUID:93100327; PMID:1281481
A:Accession: I45087
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-26 <PET>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:123686)

Query Match      25.5%; Score 28; DB 2; Length 26;
Best Local Similarity 71.4%; Pred. No. 1e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSVVIVG 9
DB 18 GSVLVVG 24

RESULT 8
E45087
cysteine proteinase homolog (clone PCR17) - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: E45087
R:Petanceska, S.; Devi, L.
J. Biol. Chem. 267, 26038-26043, 1992

```



C:Keywords: glycoprotein; plasma

Query Match 24.5%; Score 27; DB 2; Length 20;  
Best Local Similarity 40.0%; Pred. NO. 1.1e+03;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 6 VIVGRILSG 15  
|:|:|:|:  
DB 1 VVGRVYVNG 10

RESULT 14

S43635  
cytochrome-c oxidase (EC 1.9.3.1) chain VIII, hepatic - rainbow trout (fragment)  
C:Species: Oncorhynchus mykiss (rainbow trout)  
C>Date: 20-Oct-1994 #sequence\_revision 01-Nov-1996 #text\_change 18-Jul-1997  
C:Accession: S43635  
R:Freund, R.; Kadenbach, B.  
Eur. J. Biochem. 221, 1111-1116, 1994  
A:Title: Identification of tissue-specific isoforms for subunits Vb and VIIa of cytochrome c oxidase  
A:Reference number: S43624; PMID:94237150; PMID:8181469  
A:Accession: S43635  
A:Molecule type: protein  
A:Residues: 1-20 <PRE>  
A:Note: the source is designated as Salmo gairdneri  
C:Genetics:  
A:Genome: nuclear  
C:Keywords: liver; membrane-associated complex; mitochondrion; oxidoreductase

Query Match 24.5%; Score 27; DB 2; Length 20;  
Best Local Similarity 83.3%; Pred. NO. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 13 LSGRPA 18  
|:|:|:  
DB 1 LSGKPA 6

RESULT 15

S11811  
fimbrial protein - Bordetella pertussis  
C:Species: Bordetella pertussis  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 31-Oct-1997  
C:Accession: S11811  
R:Walker, M.J.; Rohde, M.; Brownlie, R.M.; Timmis, K.N.  
Mol. Microbiol. 4, 39-47, 1990  
A:Title: Engineering upstream transcriptional and translational signals of Bordetella pertussis fimbrial protein into filamentous structures.  
A:Reference number: S11811; PMID:90205626; PMID:1969607  
A:Accession: S11811  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-23 <WAL>  
C:Superfamily: type 1 fimbrial protein

Query Match 24.5%; Score 27; DB 2; Length 23;  
Best Local Similarity 55.6%; Pred. NO. 1.2e+03;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSWIVTVGRI 11  
|:|:|:|:  
DB 4 GTIVITGTI 12

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:30:53 ; Search time 16 seconds  
(without alignments)  
74.851 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110

Sequence: 1 KGSWVIVGRILSGRPAIVPRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 2073

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Query Score | Match Length | DB ID | Description |
|------------|-------------|--------------|-------|-------------|
| 1          | 29          | 26.4         | 10    | COXQ_RABIT  |
| 2          | 29          | 26.4         | 26    | SMS1_AMICA  |
| 3          | 28          | 25.5         | 15    | ACT_FINPS   |
| 4          | 28          | 25.5         | 25    | CR15_LITCE  |
| 5          | 27          | 24.5         | 20    | COXQ_ONCMY  |
| 6          | 26          | 23.6         | 26    | MEL_APIDO   |
| 7          | 26          | 23.6         | 28    | LPL_SALTY   |
| 8          | 26          | 23.6         | 28    | LPL_SALTY   |
| 9          | 25          | 22.7         | 20    | OM4V_VIBAL  |
| 10         | 25          | 22.7         | 22    | CYSP_TRIVA  |
| 11         | 25          | 22.7         | 24    | Y3KD_NEUCR  |
| 12         | 25          | 22.7         | 25    | CR11_LITSP  |
| 13         | 25          | 22.7         | 25    | CR13_LITCE  |
| 14         | 25          | 22.7         | 25    | CR1A_LITSP  |
| 15         | 24.5        | 22.3         | 30    | PSAM_CYACA  |
| 16         | 24          | 21.8         | 23    | 23KD_BAGST  |
| 17         | 24          | 21.8         | 23    | PA24_BOTAS  |
| 18         | 24          | 21.8         | 25    | BLP4_BOMOR  |
| 19         | 24          | 21.8         | 25    | CR12_LITCE  |
| 20         | 24          | 21.8         | 29    | COAL_BPI22  |
| 21         | 24          | 21.8         | 30    | AP65_CARMA  |
| 22         | 23.5        | 21.4         | 26    | VGLH_HSVF   |
| 23         | 23          | 20.9         | 15    | EPLA_MICCR  |
| 24         | 23          | 20.9         | 20    | LEC1_ARTIN  |
| 25         | 23          | 20.9         | 20    | TCM6_PACGO  |
| 26         | 23          | 20.9         | 20    | TENB_ACTIE  |
| 27         | 23          | 20.9         | 21    | LEC2_ARTIN  |
| 28         | 23          | 20.9         | 21    | ODP2_SOLTU  |
| 29         | 23          | 20.9         | 22    | HELN_HELVI  |
| 30         | 23          | 20.9         | 24    | CR16_LITXA  |
| 31         | 23          | 20.9         | 24    | CR19_LITCH  |
| 32         | 23          | 20.9         | 26    | MEL_APIEL   |
| 33         | 23          | 20.9         | 29    | PSBI_SYNVU  |

34 34 23 20.9 29 1 Y15\_BPT7 P03792 bacterioph  
35 22.5 20.5 13 1 TENA\_RANTE P56917 rana tempor  
36 22.5 20.5 16 1 H5\_COTFA P18638 coturnix co  
37 22.5 20.5 25 1 PCW1\_PACGO P82423 pachycondyl  
38 22.5 20.5 25 1 PCW2\_PACGO P82424 pachycondyl  
39 22 20.0 9 1 FARS\_ASCSU P43170 ascaris suu  
40 22 20.0 13 1 PEDI\_HYDAT P80578 hydra atten  
41 22 20.0 18 1 YPB4\_LACLC P42022 lactococcus  
42 22 20.0 20 1 CUDP\_VERCH P80406 verticilliu  
43 22 20.0 24 1 CR18\_LITCH P81251 litoria chl  
44 22 20.0 26 1 RT28\_BOVIN P82928 bos taurus  
45 22 20.0 28 1 OSTI\_CHICK P80896 gallus gall

#### ALIGNMENTS

RESULT 1  
COXQ\_RABIT  
ID COXQ\_RABIT STANDARD; PRT; 10 AA.  
AC P80336;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome c oxidase polypeptide VIII-liver/heart (EC 1.9.3.1)  
DE (fragment)  
GN COX8H.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Heart, and Liver;  
RA Freund R., Kadenbach B.;  
RL Submitted (MAR-1994) to Swiss-Prot.  
CC -!- FUNCTION: This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the terminal oxidase in mitochondrial electron transport.  
CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.  
CC -!- SIMILARITY: Belongs to the cytochrome c oxidase VIII family.  
KW Oxidoreductase; Mitochondrion.  
FT NON TER 10 10  
SQ SEQUENCE 10 AA; 1027 MW; 2C325CB40DC76338 CRC64;  
Query Match 26.4%; Score 29; DB 1; Length 10;  
Best Local Similarity 55.6%; Pred. No. 26+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 13 LSGRPAIVP 21  
:|||||  
Db 1 ISGKPARTP 9

RESULT 2  
SMS1\_AMICA  
ID SMS1\_AMICA STANDARD; PRT; 26 AA.  
AC Q9PRZ6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Somatostatin I precursor [Contains: Somatostatin-14]  
DE (fragment)  
OS Ania calva (Bowfin).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Amiiformes; Amia.  
OX NCBI\_TaxID=7924;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Pancreas;  
RX MEDLINE=9402322; PubMed=8105513;  
Wang Y., Youson J.H., Conlon J.M.;

RT "Prosomatostatin-I is processed to somatostatin-26 and somatostatin-14 in the pancreas of the bowfin, *Ambloplites caeruleus*.";  
 RL Regul. Pept. 47:33-39(1993).  
 CC -!- FUNCTION: Somatostatin inhibits the release of somatotropin.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the somatostatin family.  
 DR InterPro: IPR004250; Somatostatin.  
 DR Pfam: PF03002; Somatostatin; 1.  
 KW Cleavage on pair of basic residues; Hormone; Multigene family.  
 FT NON TER 1 1  
 FT PEPTIDE 1 26 SOMATOSTATIN-26.  
 FT PEPTIDE 13 26 SOMATOSTATIN-14.  
 FT DISULFID 15 26  
 FT SEQUENCE 26 AA; 2931 MW; 8A296DC3710552FE CRC64;  
 Query Match 26.4%; Score 29; DB 1; Length 26;  
 Best Local Similarity 55.6%; Pred. No. 4.8e+02;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 14 SCRPALVR 22  
 DB 1 SANPALVR 9  
 ID ACT\_PINPS STANDARD; PRT; 15 AA.  
 AC P81085;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Actin (Water stress responsive protein 5) (Fragment).  
 OS Pinus pinaster (Maritime pine).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OC NCBI\_TaxID=71647;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=98418576; PubMed=9747804;  
 RA Costa P., Bahman N., Frigerio J.-M., Kremer A., Plomion C.;  
 RT "Water-deficit-responsive proteins in maritime pine.";  
 RL Plant Mol. Biol. 38:587-596(1998).  
 RN [2]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=99274088; PubMed=10344291;  
 RA Costa P., Pionneau C., Bauw G., Dubos C., Bahman N., Kremer A.,  
 RA Frigerio J.-M., Plomion C.;  
 RT "Separation and characterization of needle and xylem maritime pine proteins.";  
 RL Electrophoresis 20:1098-1108(1999).  
 CC -!- FUNCTION: Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- INDUCTION: By water stress.  
 CC -!- SIMILARITY: Belongs to the actin family.  
 DR InterPro: IPR004001; Actin.  
 DR InterPro: IPR004000; Actin like.  
 DR PROSITE; PS00406; ACTINS\_1; PARTIAL.  
 DR PROSITE; PS01132; ACTINS\_ACT LIKE; PARTIAL.  
 DR PROSITE; PS00432; ACTINS\_2; PARTIAL.  
 KW Structural protein.  
 FT NON TER 1 1  
 FT NON TER 15 15  
 FT SEQUENCE 15 AA; 1574 MW; 4712F0146D5B4DE3 CRC64;  
 Query Match 25.5%; Score 28; DB 1; Length 15;  
 Best Local Similarity 55.6%; Pred. No. 4.1e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 7 IVGRILSG 15  
 ID ACT\_PINPS STANDARD; PRT; 15 AA.  
 AC P81085;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Actin (Water stress responsive protein 5) (Fragment).  
 OS Pinus pinaster (Maritime pine).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OC NCBI\_TaxID=71647;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=98418576; PubMed=9747804;  
 RA Costa P., Bahman N., Frigerio J.-M., Kremer A., Plomion C.;  
 RT "Water-deficit-responsive proteins in maritime pine.";  
 RL Plant Mol. Biol. 38:587-596(1998).  
 RN [2]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=99274088; PubMed=10344291;  
 RA Costa P., Pionneau C., Bauw G., Dubos C., Bahman N., Kremer A.,  
 RA Frigerio J.-M., Plomion C.;  
 RT "Separation and characterization of needle and xylem maritime pine proteins.";  
 RL Electrophoresis 20:1098-1108(1999).  
 CC -!- FUNCTION: Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- INDUCTION: By water stress.  
 CC -!- SIMILARITY: Belongs to the actin family.  
 DR InterPro: IPR004001; Actin.  
 DR InterPro: IPR004000; Actin like.  
 DR PROSITE; PS00406; ACTINS\_1; PARTIAL.  
 DR PROSITE; PS01132; ACTINS\_ACT LIKE; PARTIAL.  
 DR PROSITE; PS00432; ACTINS\_2; PARTIAL.  
 KW Structural protein.  
 FT NON TER 1 1  
 FT NON TER 15 15  
 FT SEQUENCE 15 AA; 1574 MW; 4712F0146D5B4DE3 CRC64;  
 Query Match 25.5%; Score 28; DB 1; Length 15;  
 Best Local Similarity 55.6%; Pred. No. 4.1e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 7 IVGRILSG 15

Db 2 LGYNVLSG 10  
 ID CR15\_LITCE STANDARD; PRT; 25 AA.  
 AC P56230;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Caerin 1.5.  
 OS Litoria caerulea (Green tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylidae;  
 OC Pelodyadinae; Litoria.  
 OC NCBI\_TaxID=50344;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Parotoid gland;  
 RA Stone D.J.M., Waugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "Peptides from Australian frogs. The structures of the caerins from Litoria caerulea.";  
 RL J. Chem. Res. 138:910-936(1993).  
 CC -!- FUNCTION: Antibacterial peptide, that adopts an alpha helical conformation which can disrupt bacterial membranes. Each caerin displays a different antimicrobial specificity.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Secreted by the skin parotoid and/or rostral glands.  
 CC -!- DOMAIN: Contains two amphipathic alpha helix regions separated by a region of less-defined helicity and greater flexibility (By similarity).  
 CC -!- MASS SPECTROMETRY: MW=2610; METHOD=FAB.  
 CC Amphibian defense peptide; Antibiotic; Amidation.  
 FT MOD RES 25 25  
 FT SEQUENCE 25 AA; 2613 MW; 0FF5A464EA0BBE12 CRC64;  
 Query Match 25.5%; Score 28; DB 1; Length 25;  
 Best Local Similarity 26.3%; Pred. No. 6.6e+02;  
 Matches 5; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 3 GSVIVIGRIILSGRPAIVP 21  
 DB 1 GLSLVLSGVVKKVIPHVP 19  
 ID COXQ\_ONCMY STANDARD; PRT; 20 AA.  
 AC P80335;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome c oxidase polypeptide VIII-liver/heart (EC 1.9.3.1) (Fragment).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OC NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Heart, and Liver;  
 RX MEDLINE=94217150; PubMed=8181469;  
 RA Freund R., Kadenbach B.;  
 RT "Identification of tissue-specific isoforms for subunits Vb and VIIa of cytochrome c oxidase isolated from rainbow trout.";  
 RL Eur. J. Biochem. 221:1111-1116(1994).  
 CC -!- FUNCTION: This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the terminal oxidase in mitochondrial electron transport.  
 CC -!- CATALYTIC ACTIVITY: 4 ferrocyanochrome c + O(2) = 4 ferricytochrome

```

CC C + 2 H(2)O.
CC -!- SIMILARITY: Belongs to the cytochrome c oxidase VIII family.
KW PIR; S43635; S43635.
DR Oxidoreductase; Mitochondrion.
FT NON TER 20
SQ SEQUENCE 20 AA; 2048 MW; 4977B9DE80562C58 CRC64;

Query Match 24.5%; Score 27; DB 1; Length 20;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 13 LSGRPA 18
DB 1 LSGRPA 6

RESULT 6
MEL_APIDO
ID_MEL_APIDO STANDARD; PRT; 26 AA.
AC P01502;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Melittin.
GN MELT.
OS Apis dorsata (Giant honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apoidea; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7462;
RN [1]
RP SEQUENCE.
RX MEDLINE=75168194; PubMed=1093875;
RA Kreil G.;
RA "The structure of Apis dorsata melittin: phylogenetic relationships
between honeybees as deduced from sequence data.";
RL FEBS Lett. 54:100-102(1975).
CC -!- FUNCTION: Main toxin of bee venom with strong hemolytic activity.
CC Integrates into cell membranes and has multiple effects, probably,
CC as a result of its interaction with negatively charged
CC phospholipids. It inhibits well known transport pumps such as the
CC Na(+)-K(+)-ATPase and the H(+)-K(+)-ATPase. Increases the
CC permeability of cell membranes to ions, particularly Na+ and
CC indirectly Ca(2+), because of the Na(+)-Ca(2+)-exchange.
CC -!- SUBUNIT: Monomer and homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland (Probable).
CC -!- SIMILARITY: Belongs to the melittin family.
CC -!- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 12 of July 2001;
CC WWW="http://www.expasy.org/spotlight/articles/splt012.html".
DR PIR; A01763; MEHBCD.
DR HSP; P01501; 1BH1.
DR InterPro; IPR002116; Melittin.
DR Pfam; PF01372; Melittin; 1.
DR ProDom; PD014636; Melittin; 1.
KW Cytolysis; Hemolysis; Toxin; Amidation.
FT MOD RES 26
SQ SEQUENCE 26 AA; 2848 MW; F1DA8F92514EF01C CRC64;

Query Match 23.6%; Score 26; DB 1; Length 26;
Best Local Similarity 27.3%; Pred. No. 1.4e+03;
Matches 3; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 10 RIILSGRPAIV 20
DB 7 KVLSTGLPALI 17

RESULT 7
LPL_SALTY
ID_LPL_SALTY STANDARD; PRT; 28 AA.
AC Q825H9;

Query Match 23.6%; Score 26; DB 1; Length 28;
Best Local Similarity 30.8%; Pred. No. 1.5e+03;
Matches 4; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 VVIVGRILSGRP 17
DB 10 LLLNAFIVGRP 22

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leu operon leader peptide.
GN LEUL OR STY0133 OR T0118.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]_TaxID=601;
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner P.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Involved in control of the biosynthesis of leucine (By
similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; AL627265; CAD01271.1; --
DR EMBL; AF016834; AAO67850.1; --
KW Leucine biosynthesis; Leader peptide; Complete proteome.
SQ SEQUENCE 28 AA; 3129 MW; 6381476044552145 CRC64;

Query Match 23.6%; Score 26; DB 1; Length 28;
Best Local Similarity 30.8%; Pred. No. 1.5e+03;
Matches 4; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 VVIVGRILSGRP 17
DB 10 LLLNAFIVGRP 22

```

```

RC STRAIN=LT2;
RX MEDLINE=80056609; PubMed=388423;
RA Gemmell R.M., Wessler S.R., Keller E.B., Calvo J.M.;
RT "Leu operon of Salmonella typhimurium is controlled by an attenuation
RT mechanism.";
RL Proc. Natl. Acad. Sci. U.S.A. 76:4941-4945 (1979).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
RC -!- FUNCTION: Involved in control of the biosynthesis of leucine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00059; -; NOT ANNOTATED CDS.
DR EMBL; AF008699; AAL13078.1; -.
DR PIR; A03598; LFEHLT.
DR StyGene; SG10199; leul.
KW Leucine biosynthesis; Leader peptide; Complete proteome.
SQ SEQUENCE 28 AA; 3103 MW; 6391566144552145 CRC64;
-----
Query Match 23.6%; Score 26; DB 1; Length 28;
Best Local Similarity 30.8%; Pred. No. 1.5e+03;
Matches 4; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 VVIVGRILSGRP 17
Db : : : : :
10 LLLNFAFVGRGP 22

RESULT 9
OM4V VIBAL STANDARD; PRT; 20 AA.
AC P83129;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein 40Va (Omp40Va) (Fragment).
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE.
RC STRAIN=NCIMB 1903T;
RA Onji M., Hirabayashi J., Suzuki S.;
RT "Characterization of major outer membrane proteins of Vibrio
RT alginolyticus and the stability against proteases.";
RL Microbes Environ. 0:0-0 (2002).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the Gram-negative porin family.
KW Transmembrane; Porin; Outer membrane.
FT NON TER 20
SQ SEQUENCE 20 AA; 2229 MW; 4FA95CB24FEC3CDB CRC64;
-----
Query Match 22.7%; Score 25; DB 1; Length 20;
Best Local Similarity 38.9%; Pred. No. 1.5e+03;
Matches 7; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

QY 5 VVIVGRILSGRP 17
Db : : : : :
10 LLLNFAFVGRGP 22

RESULT 9
OM4V VIBAL STANDARD; PRT; 20 AA.
AC P83129;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein 40Va (Omp40Va) (Fragment).
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE.
RC STRAIN=NCIMB 1903T;
RA Onji M., Hirabayashi J., Suzuki S.;
RT "Characterization of major outer membrane proteins of Vibrio
RT alginolyticus and the stability against proteases.";
RL Microbes Environ. 0:0-0 (2002).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the Gram-negative porin family.
KW Transmembrane; Porin; Outer membrane.
FT NON TER 20
SQ SEQUENCE 20 AA; 2229 MW; 4FA95CB24FEC3CDB CRC64;
-----
Query Match 22.7%; Score 25; DB 1; Length 22;
Best Local Similarity 40.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 KKGSVVIVGR 10
Db : : : : :
9 KKGAVNVIXK 18

RESULT 11
Y3KD NEUCR STANDARD; PRT; 24 AA.
AC P22702;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 2.8 kDa protein in ARG-2 5' region.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90293033; PubMed=2141606;
RA Orbach M.J., Sachs M.S., Yanofsky C.;
RT "The Neurospora crassa arg-2 locus. Structure and expression of the
RT gene encoding the small subunit of arginine-specific carbamoyl
RT phosphate synthetase."
RL J. Biol. Chem. 265:10981-10987 (1990).
CC -!- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE PRODUCTION OF
CC ARGININE-SPECIFIC CARBAMOYL PHOSPHATE SYNTHETASE.
CC -!- SIMILARITY: TO YEAST CPA1 LEADER PEPTIDE (AC P08521).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

QY 1 KKGSVVIVGRILSGRPA 18
Db : : : : :
7 EEGSVDFYQQL----RPA 20

RESULT 10
CYSP TRIVA STANDARD; PRT; 22 AA.
ID CYSP TRIVA STANDARD; PRT; 22 AA.
AC P33404;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cysteine proteinase (EC 3.4.22.-) (Fragment).
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
OC Trichomonadinae; Trichomonas.
OX NCBI_TaxID=5722;
RN [1]
RP SEQUENCE.
RX MEDLINE=93307628; PubMed=8319888;
RA Irvine J.W., Coombs G.H., North M.J.;
RT "Purification of cysteine proteinases from trichomonads using
RT bacitracin-Sepharose."
RL FEMS Microbiol. Lett. 110:113-120 (1993).
CC -!- SIMILARITY: Belongs to peptidase family C1.
DR MEROPS; C01.082; -.
DR InterPro; IPR000169; SHProt acsite.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
KW Hydrolase; Thiol protease.
FT UNSURE 1
FT NON TER 22
SQ SEQUENCE 22 AA; 2398 MW; 03E40FD86661ACCB CRC64;
-----
Query Match 22.7%; Score 25; DB 1; Length 22;
Best Local Similarity 40.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 KKGSVVIVGR 10
Db : : : : :
9 KKGAVNVIXK 18

RESULT 11
Y3KD NEUCR STANDARD; PRT; 24 AA.
AC P22702;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 2.8 kDa protein in ARG-2 5' region.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90293033; PubMed=2141606;
RA Orbach M.J., Sachs M.S., Yanofsky C.;
RT "The Neurospora crassa arg-2 locus. Structure and expression of the
RT gene encoding the small subunit of arginine-specific carbamoyl
RT phosphate synthetase."
RL J. Biol. Chem. 265:10981-10987 (1990).
CC -!- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE PRODUCTION OF
CC ARGININE-SPECIFIC CARBAMOYL PHOSPHATE SYNTHETASE.
CC -!- SIMILARITY: TO YEAST CPA1 LEADER PEPTIDE (AC P08521).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMIL; J05512; AAA33610.1; --  
 DR PIR; B4224; B4224.  
 KW Hypothetical protein.  
 SQ SEQUENCE 24 AA; 2779 MW; 8FF3F527EAD244C CRC64;

Query Match 22.7%; Score 25; DB 1; Length 24;  
 Best Local Similarity 42.9%; Pred. No. 1.8e+03;  
 Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LSGRPAI 19  
 : : : : :  
 Db 1 MNGRPSV 7

## RESULT 12

CR11\_LITSP STANDARD; PRT; 25 AA.  
 AC P56226;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Caerin 1.1  
 OS Litoria splendida (Magnificent tree frog),  
 OS Litoria gilleni (Centralian tree frog), and  
 OS Litoria caerulea (Green tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae;  
 OC Pelodyadinae; Litoria  
 OC NCBI\_TaxID=30345, 39405, 30344;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY OF 1.1; 1.1.1 AND 1.1.2.  
 RC SPECIES=L.splendida; TISSUE=Parotoid gland;  
 RA Stone D.J.M., Waugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "Peptides from Australian frogs. Structures of the caerins and  
 RT caeridin 1 from Litoria splendida".  
 RL J. Chem. Soc. Perkin Trans. 1:13173-13178 (1992).  
 RN [2]  
 RP SEQUENCE, AND MASS SPECTROMETRY OF 1.1; 1.1.1; 1.1.2 AND 1.1.4.  
 RC SPECIES=L.caerulea; TISSUE=Parotoid gland;  
 RA Stone D.J.M., Waugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "Peptides from Australian frogs. The structures of the caerins from  
 RT Litoria caerulea".  
 RL J. Chem. Res. 138:910-936 (1993).  
 RN [3]  
 RP SEQUENCE, AND MASS SPECTROMETRY OF 1.1; 1.1.1; 1.1.3; 1.1.5 TO 1.1.8.  
 RC SPECIES=L.gilleni; TISSUE=Parotoid gland;  
 RA Waugh R.J., Stone D.J.M., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "Peptides from Australian frogs. The structures of the caerins and  
 RT caeridins from Litoria gilleni".  
 RL J. Chem. Res. 139:937-961 (1993).  
 RN [4]  
 RP FUNCTION, AND STRUCTURE BY NMR.  
 RC SPECIES=L.splendida;  
 RX MEDLINE=97409981; PubMed=9266696;  
 RA Wong H., Bowie J.H., Carver J.A.;  
 RT "The solution structure and activity of caerin 1.1, an antimicrobial  
 RT peptide from the Australian green tree frog, Litoria splendida".  
 RL Eur. J. Biochem. 247:545-557 (1997).  
 CC -!- FUNCTION: Antibacterial and antiviral peptides that adopt an alpha  
 CC helical conformation which can disrupt bacterial membranes. Each  
 CC caerin displays a different antimicrobial specificity.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Secreted by the skin parotoid and/or rostral  
 CC glands.  
 CC -!- DOMAIN: Contains two amphipathic alpha helix regions separated by  
 CC a region of less-defined helicity and greater flexibility.  
 CC -!- PTM: THE MAJOR PRODUCT IS CAERIN 1.1; IN ADDITION, DIFFERENT  
 CC PEPTIDES ARE PRODUCED THAT ARE MISSING SOME AMINO ACID RESIDUES AT  
 CC THE N-TERMINUS OR C-TERMINUS. ALL ISOFORMS ARE NOT REPRESENTED IN

CC EACH SPECIES. CAERIN 1.1.1 AND CAERIN 1.1.4 ARE INACTIVE.  
 CC -!- MASS SPECTROMETRY: MW=2582; METHOD=FAB; RANGE=1-25.  
 CC -!- MASS SPECTROMETRY: MW=2412; METHOD=FAB; RANGE=3-25.  
 CC -!- MASS SPECTROMETRY: MW=2299; METHOD=FAB; RANGE=4-25.  
 CC -!- MASS SPECTROMETRY: MW=1421; METHOD=FAB; RANGE=13-25.  
 CC -!- MASS SPECTROMETRY: MW=2333; METHOD=FAB; RANGE=1-23.  
 CC -!- MASS SPECTROMETRY: MW=1626; METHOD=FAB; RANGE=1-16.  
 CC -!- MASS SPECTROMETRY: MW=1489; METHOD=FAB; RANGE=1-15.  
 CC -!- MASS SPECTROMETRY: MW=1180; METHOD=FAB; RANGE=1-12.  
 CC -!- MASS SPECTROMETRY: MW=915; METHOD=FAB; RANGE=1-10.  
 KW Amphibian defense peptide; Antibiotic; Antiviral; Amidation.  
 FT PEPTIDE 1 25  
 FT CAERIN 1.1.1  
 FT PEPTIDE 3 25  
 FT CAERIN 1.1.2  
 FT PEPTIDE 4 25  
 FT CAERIN 1.1.3  
 FT PEPTIDE 13 25  
 FT CAERIN 1.1.4  
 FT PEPTIDE 1 16  
 FT CAERIN 1.1.5  
 FT PEPTIDE 1 15  
 FT CAERIN 1.1.6  
 FT PEPTIDE 1 12  
 FT CAERIN 1.1.7  
 FT PEPTIDE 1 10  
 FT CAERIN 1.1.8  
 FT MOD RES 25  
 FT AMIDATION.  
 SQ SEQUENCE 25 AA; 2585 MW; D8A5A460BB0EBE00 CRC64;

Query Match 22.7%; Score 25; DB 1; Length 25;  
 Best Local Similarity 26.3%; Pred. No. 1.9e+03;  
 Matches 5; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GSVIVIGRIILSGRPAIVP 21  
 : : : : :  
 Db 1 GLLSVLGSVAQHVLPHPVP 19

## RESULT 13

CR13\_LITCE STANDARD; PRT; 25 AA.  
 ID CR13\_LITCE  
 AC P56228;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Caerin 1.3.  
 OS Litoria caerulea (Green tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae;  
 OC Pelodyadinae; Litoria.  
 OC NCBI\_TaxID=30344;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Parotoid gland;  
 RA Stone D.J.M., Waugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "Peptides from Australian frogs. The structures of the caerins from  
 RT Litoria caerulea".  
 RL J. Chem. Res. 138:910-936 (1993).  
 CC -!- FUNCTION: Antibacterial peptide, that adopts an alpha helical  
 CC conformation which can disrupt bacterial membranes. Each caerin  
 CC displays a different antimicrobial specificity.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Secreted by the skin parotoid and/or rostral  
 CC glands.  
 CC -!- DOMAIN: Contains two amphipathic alpha helix regions separated by  
 CC a region of less-defined helicity and greater flexibility (By  
 CC similarity).  
 CC -!- MASS SPECTROMETRY: MW=2582; METHOD=FAB.  
 KW Amphibian defense peptide; Antibiotic; Amidation.  
 FT MOD RES 25  
 FT AMIDATION.  
 SQ SEQUENCE 25 AA; 2585 MW; D8A5A460BB0E2F2 CRC64;

Query Match 22.7%; Score 25; DB 1; Length 25;  
 Best Local Similarity 26.3%; Pred. No. 1.9e+03;  
 Matches 5; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GSVIVIGRIILSGRPAIVP 21  
 : : : : :  
 Db 1 GLLSVLGSVAQHVLPHPVP 19

## RESULT 14

CRIA\_LITSP STANDARD; PRT; 25 AA.  
 ID P82104;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Caerin 1.10.  
 OS Litoria splendida (Magnificent tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylinae;  
 OC Pelodyadinae; Litoria.  
 OX NCBI\_TaxID=30345;  
 RN [1]  
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
 RC TISSUE=Skin secretion;  
 RX MEDLINE=20069371; PubMed=10601876;  
 RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
 RT "Differences in the skin peptides of the male and female Australian  
 tree frog Litoria splendida. The discovery of the aquatic male sex  
 pheromone splendipherin, together with Phe8 caerulein and the  
 antibiotic peptide caerin 1.10.";  
 RT antibiotic peptide caerin 1.10.";  
 RL Eur. J. Biochem. 267:269-275(2000).  
 CC -!- FUNCTION: Antibacterial peptide with wide spectrum of activity.  
 CC Active against L.lactis, L.linnocua, M.luteus, S.uberis and less  
 CC against P.multocida and S.epidermidis.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Skin.  
 CC -!- MASS SPECTROMETRY: MW=2573; METHOD=Electrospray.  
 CC -!- MISCELLANEOUS: Caerin 1.10 is not present in the glandular  
 CC secretion of female L.splendida.  
 CC Amphibian defense peptide; Antibiotic; Amidation.  
 KW MOD\_RES 25 25 AMIDATION.  
 FT MOD\_RES 25 25  
 SQ SEQUENCE 25 AA; 2576 MW; DBA77460BB0EBE00 CRC64;

Query Match 22.7%; Score 25; DB 1; Length 25;  
 Best Local Similarity 26.3%; Pred. No. 1.9e+03;  
 Matches 5; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 3 GSWIVIGRIILSGRPAIVP 21  
 | : : : : :  
 Db 1 GLLSVLGSVAKHVLPHVVP 19

## RESULT 15

PSAM\_CYACA STANDARD; PRT; 30 AA.  
 ID Q9TLX5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Photosystem I reaction centre subunit XII (PSI-M).  
 GN PSAM.  
 OS Cyanidium caldarium.  
 OG Chloroplast.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;  
 OC Cyanidium.  
 OX NCBI\_TaxID=2771;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RK-1;  
 RX MEDLINE=20496959; PubMed=11040290;  
 RA Gloeckner G., Rosenthal A., Valentin K.-U.;  
 RT "The structure and gene repertoire of an ancient red algal plastid  
 genome.";  
 RL J. Mol. Evol. 51:382-390(2000).  
 CC -!- SIMILARITY: Belongs to the psam family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC -----  
 DR EMBL; AF022186; AAF12951.1; -- Chloroplast.  
 KW Photosystem I; Photosynthesis; ECEFB5AF44D6F6 CRC64;  
 SQ SEQUENCE 30 AA; 3236 MW; ECEFB5AF44D6F6 CRC64;

Query Match 22.3%; Score 24.5; DB 1; Length 30;  
 Best Local Similarity 38.9%; Pred. No. 2.6e+03;  
 Matches 7; Conservative 3; Mismatches 3; Indels 5; Gaps 1;  
 QY 4 SVVIVGRIILSGRPAIVP 21  
 | : : : : :  
 Db 5 SQVFIGLVI-----ALVP 17

Search completed: July 6, 2004, 09:42:30  
 Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:34:38 ; Search time 58 Seconds  
(without alignments)  
125.119 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110

Sequence: 1 KGSVWVIGRIILSGRAIVPRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 18020

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mbc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_virus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 36    | 32.7        | 19     | 16 Q9K8M8 | Q9K8M8 bacillus ha  |
| 2          | 35    | 31.8        | 18     | 4 Q9HBD9  | Q9HBD9 homo sapien  |
| 3          | 32.5  | 29.5        | 30     | 11 Q9CWB1 | Q9CWB1 mus musculus |
| 4          | 32    | 29.1        | 25     | 12 Q9QP62 | Q9QP62 hepatitis c  |
| 5          | 31    | 28.2        | 20     | 1 Q3UWL5  | Q3UWL5 methanococc  |
| 6          | 31    | 28.2        | 20     | 1 Q9DD39  | Q9DD39 nipponia ni  |
| 7          | 31    | 28.2        | 22     | 13 Q57692 | Q57692 ctis tarda.  |
| 8          | 31    | 28.2        | 22     | 13 Q919P3 | Q919P3 ara militar  |
| 9          | 31    | 28.2        | 23     | 13 Q9PR50 | Q9PR50 gadus morhu  |
| 10         | 30    | 27.3        | 21     | 10 Q41181 | Q41181 nicotiana t  |
| 11         | 30    | 27.3        | 29     | 8 Q35738  | Q35738 rattus norv  |
| 12         | 30    | 27.3        | 29     | 12 Q84069 | Q84069 influenzavi  |
| 13         | 30    | 27.3        | 30     | 2 Q54460  | Q54460 serratia ma  |
| 14         | 30    | 27.3        | 30     | 16 Q7VHK2 | Q7VHK2 helicobacte  |
| 15         | 29    | 26.4        | 16     | 12 Q83960 | Q83960 influenzavi  |
| 16         | 29    | 26.4        | 16     | 12 Q84055 | Q84055 influenzavi  |

|    |      |      |    |    |        |                    |
|----|------|------|----|----|--------|--------------------|
| 17 | 29   | 26.4 | 25 | 16 | Q26056 | O26056 helicobacte |
| 18 | 29   | 26.4 | 27 | 5  | Q76549 | Q76549 haliotis fu |
| 19 | 29   | 26.4 | 27 | 11 | Q99KX5 | Q99KX5 mus musculu |
| 20 | 29   | 26.4 | 27 | 13 | O57554 | O57554 lampetra pl |
| 21 | 29   | 26.4 | 29 | 11 | Q8CGM8 | Q8CGM8 mus musculu |
| 22 | 29   | 26.4 | 30 | 8  | O98712 | O98712 euglena ana |
| 23 | 27.5 | 25.0 | 17 | 10 | O65345 | O65345 gossypium h |
| 24 | 27.5 | 25.0 | 30 | 16 | Q8U545 | Q8U545 agrobacteri |
| 25 | 27   | 24.5 | 18 | 4  | Q8N303 | Q8N303 homo sapien |
| 26 | 27   | 24.5 | 20 | 4  | Q8TE41 | Q8TE41 homo sapien |
| 27 | 27   | 24.5 | 20 | 13 | Q9PS35 | Q9PS35 carcharinu  |
| 28 | 27   | 24.5 | 21 | 11 | Q9QV63 | Q9QV63 rattus sp.  |
| 29 | 27   | 24.5 | 27 | 12 | Q91L56 | Q91L56 hepatitis c |
| 30 | 27   | 24.5 | 28 | 8  | Q9T4F8 | Q9T4F8 meleagris g |
| 31 | 27   | 24.5 | 28 | 8  | Q9T7L2 | Q9T7L2 meleagris g |
| 32 | 27   | 24.5 | 28 | 8  | Q9T4F7 | Q9T4F7 meleagris g |
| 33 | 27   | 24.5 | 28 | 8  | Q9T7L1 | Q9T7L1 meleagris g |
| 34 | 27   | 24.5 | 29 | 4  | Q96BS8 | Q96BS8 homo sapien |
| 35 | 27   | 24.5 | 29 | 9  | Q7Y5C6 | Q7Y5C6 bacterioph  |
| 36 | 27   | 24.5 | 30 | 6  | Q9TQU1 | Q9TQU1 equus cabal |
| 37 | 27   | 24.5 | 30 | 6  | Q9TQU2 | Q9TQU2 equus cabal |
| 38 | 26.5 | 24.1 | 26 | 5  | P90723 | P90723 berce ovata |
| 39 | 26   | 23.6 | 14 | 10 | Q9FUX5 | Q9FUX5 symphoricar |
| 40 | 26   | 23.6 | 18 | 12 | Q84129 | Q84129 influenzavi |
| 41 | 26   | 23.6 | 21 | 5  | Q9TWT7 | Q9TWT7 plasmodium  |
| 42 | 26   | 23.6 | 21 | 6  | Q95XS4 | Q95XS4 ovis aries  |
| 43 | 26   | 23.6 | 22 | 6  | Q95KU7 | Q95KU7 bos taurus  |
| 44 | 26   | 23.6 | 24 | 13 | Q9PR51 | Q9PR51 gadus morhu |
| 45 | 26   | 23.6 | 27 | 15 | Q9ENL6 | Q9ENL6 human immun |

#### ALIGNMENTS

RESULT 1

Q9K8M8 PRELIMINARY; PRT; 19 AA.  
 ID Q9K8M8  
 AC Q9K8M8;  
 DT 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Hypothetical protein BH2977.  
 GN BH2977.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 CX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL, AF001517, BAB06696.1; -;  
 DR FIK; AB4022; AB4022.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 19 AA; 2165 MW; 8CD3EAFAD137871D CRC64;

Query Match 32.7%; Score 36; DB 16; Length 19;

Best Local Similarity 66.7%; Pred. No. 1.8e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KGSVWVVG 9

||||:|

6 KGSIIIFVG 14

RESULT 2

Q9HBD9 PRELIMINARY; PRT; 18 AA.

ID Q9HBD9

```

AC Q9HBD9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ATP7B (Fragment).
GN ATP7B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu Z.Y., Wang N., MuRong S.X.;
RT "Ala874Val (2621C-T) missense mutation of ATP7B gene in Chinese.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF254561; RAG27538.1; -.
FT NON_TER 1 1
FT VARIANT 15 15 V -> A.
FT NON_TER 18 18
SQ SEQUENCE 18 AA; 1814 MW; CEDCEA2D27041AFA CRC64;

Query Match 31.8%; Score 35; DB 4; Length 18;
Best Local Similarity 72.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRI 11
Db 8 KPGSTVIVGSI 18

RESULT 3
Q9CWB1 PRELIMINARY; PRT; 30 AA.
ID Q9CWB1
AC Q9CWB1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Pyruvate kinase 3 (Fragment).
GN PKM2 OR PK3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL; AK002341; BAB2025.1; -.
DR HSSP; F11974; 1A49.
DR MGD; MGI:97591; Pkm2.
DR GO; GO:0004743; F:pyruvate kinase activity; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001697; Pyruvate_kinase.

DR Pfam; PF02887; PK_C; 1.
DR ProDom; PD001009; Pyruvate_kinase; 1.
FT NON_TER 1 1
SQ SEQUENCE 30 AA; 3306 MW; BF0DSA1EC70809DF CRC64;

Query Match 29.5%; Score 32.5; DB 11; Length 30;
Best Local Similarity 61.1%; Pred. No. 9.7e+02;
Matches 11; Conservative 1; Mismatches 1; Indels 5; Gaps 2;

QY 1 KKGSVVIVGRIILSG-RP 17
Db 3 KKGDVVIV----LTGWRP 16

RESULT 4
Q9QP62 PRELIMINARY; PRT; 25 AA.
ID Q9QP62
AC Q9QP62
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Truncated NS2.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Leng Y., Chen X.Z., Du Y., Wang H.T.;
RT "Cloning of NS2 cDNA of Hepatitis C virus (HCV) and complexity of NS2
RT quasiespecies in a HCV carrier.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF175472; AAD50790.1; -.
SQ SEQUENCE 25 AA; 2637 MW; 5684D08FA4DAC202 CRC64;

Query Match 29.1%; Score 32; DB 12; Length 25;
Best Local Similarity 46.7%; Pred. No. 9.6e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 GSVVIVGRIILSGRP 17
Db 10 GGVVFGVLVFLTLP 24

RESULT 5
Q9UWL5 PRELIMINARY; PRT; 20 AA.
ID Q9UWL5
AC Q9UWL5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Membrane-associated ATPase alpha subunit (Fragment).
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2188;
RN [1]
RP SEQUENCE.
RA Chen W., Konisky J.;
RT "Characterization of a membrane-associated ATPase from Methanococcus
RT voltae, a methanogenic member of the Archaea.";
RL J. Bacteriol. 175:5677-5682 (1993).
SQ SEQUENCE 20 AA; 1965 MW; 8251D2DCB4483FA5 CRC64;

Query Match 28.2%; Score 31; DB 1; Length 20;
Best Local Similarity 42.9%; Pred. No. 1.1e+03;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 7 IVGRIILSGRPAIV 20
Db 1 VVGKIIRIAGPVVV 14

```

```
RESULT 6
Q9DD39          PRELIMINARY;      PRT;      21 AA.
ID Q9DD39
AC Q9DD39
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Non-W chromodomain helicase DNA binding protein (W-linked chromodomain
GN NON-W CHD OR W-LINKED CHD.
OS Nipponia nippon.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Ciconiiformes; Threskiornithidae;
OC Nipponia.
OX NCBI_TaxID=128390;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai Y., Ishii S., Kikuchi M.;
RT "Nipponia nippon non-W CHD gene.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arai Y., Ishii S., Kikuchi M.;
RT "Nipponia nippon CHD gene W-linked.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049444; BAB15805.1; -
DR EMBL; AB049443; BAB15804.1; -
DR GO; GO:0004386; F:helicase activity; IEA.
KW Helicase.
FT NON_TER 1 1
FT NON_TER 21 21
SQ SEQUENCE 21 AA; 2588 MW; C8E13B3B1AD353E3 CRC64;

Query Match 28.2%; Score 31; DB 13; Length 21;
Best Local Similarity 62.5%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GRPAIVPR 22
Db 11 GRPTIPR 18

RESULT 7
O57692          PRELIMINARY;      PRT;      22 AA.
ID O57692
AC O57692
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DE CHROMO-helicase-DNA-binding ON A NON-W chromosome protein (Fragment).
GN CHD OR CHD-W.
OS Otis tarda.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Gruiformes; Otididae; Otis.
OX NCBI_TaxID=73107;
RN [1]
RP SEQUENCE FROM N.A.
RA Pitra C.;
RT STRAIN=016, and D16;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047391; AAC04869.1; -
DR EMBL; AF047390; AAC04868.1; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
KW DNA-binding; Helicase.
FT NON_TER 1 1
FT NON_TER 22 22
SQ SEQUENCE 22 AA; 2675 MW; C8E13B3B1ABA5573 CRC64;

Query Match 28.2%; Score 31; DB 13; Length 22;
Best Local Similarity 62.5%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 8
Q919P3          PRELIMINARY;      PRT;      22 AA.
ID Q919P3
AC Q919P3
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Chromo-helicase-DNA binding protein (Fragment).
OS Ara militaris.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Psittaciformes; Psittacidae; Ara.
OX NCBI_TaxID=57237;
RN [1]
RP SEQUENCE FROM N.A.
RA Leal-Garza C.H., Bermudez-Humaran L.G., Jaramillo-Rangel G.,
RA Riojas-Valdes V., Montes de Oca-Luna R.;
RT "Molecular sexing of monomorphic endangered Ara birds.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229847; AAF42915.1; -
DR GO; GO:0004386; F:helicase activity; IEA.
KW Helicase.
FT NON_TER 1 1
FT NON_TER 22 22
SQ SEQUENCE 22 AA; 2703 MW; C8E13B3B1AA56873 CRC64;

Query Match 28.2%; Score 31; DB 13; Length 22;
Best Local Similarity 62.5%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GRPAIVPR 22
Db 12 GRPTIPR 19

RESULT 9
Q9PRSO          PRELIMINARY;      PRT;      23 AA.
ID Q9PRSO
AC Q9PRSO
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Chymotrypsin isoenzyme CHT2 (Fragment).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE.
RX MEDLINE=96043258; PubMed=7584866;
RA Raai A.J., Flengsrud R., Sletten K.;
RT "Chymotrypsin isoenzymes in Atlantic cod; differences in kinetics and
RT substrate specificity.";
RL Comp. Biochem. Physiol. 112B:393-398(1995).
FT NON_TER 1 1
FT NON_TER 23 23
SQ SEQUENCE 23 AA; 2356 MW; A5EC2A7B73EB3F5E CRC64;

Query Match 28.2%; Score 31; DB 13; Length 23;
Best Local Similarity 85.7%; Pred. No. 1.2e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 15 GRPAIVPR 21
Db 2 GRPAISP 8
```

```

RESULT 10
Q41181 ID Q41181 PRELIMINARY; PRT; 21 AA.
AC Q41181;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Basic beta-1,3-glucanase (Fragment).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92404740; PubMed=2152343;
RA Neale A.D., Wahlthner J.A., Lund M., Bennett H.T., Kelly A.,
RA Meeks-Wagner D.R., Peacock W.J., Dennis E.S.; and
RT tobacco explants during flower formation."
RL Plant Cell 2:673-684(1990).
DR EMBL; S44871; AAB23377.1; -.
DR FIR; PQ0145; PQ0145.
DR GO; GO:0004553; R:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000490; Glyco_hydro_17.
DR Pfam; PF00332; Glyco_hydro_17; 1.
FT NON_TER 1
SQ SEQUENCE 21 AA; 2481 MW; 5482B9F6154A7547 CRC64;

Query Match 27.3%; Score 30; DB 10; Length 21;
Best Local Similarity 71.4%; Pred. No. 1.6e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 17 PAIVPRR 23
DB 13 PSIVPKR 19

RESULT 11
Q35738 ID Q35738 PRELIMINARY; PRT; 29 AA.
AC Q35738;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rat mitochondrial type A and B DNA, HindIII (Fragment).
OS Rattus norvegicus (Rat).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81101062; PubMed=6256736;
RA Castora F.J., Arnhem N., Simpson M.V.;
RT "Mitochondrial DNA polymorphism: evidence that variants detected by
RT restriction enzymes differ in nucleotide sequence rather than in
RT methylation."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6415-6419(1980).
DR EMBL; M10477; AA59710.1; -.
DR FIR; I58970; I58970.
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 1
SQ SEQUENCE 29 AA; 3315 MW; 8D9A3773AFB6A2AB CRC64;

Query Match 27.3%; Score 30; DB 8; Length 29;
Best Local Similarity 44.4%; Pred. No. 2.2e+03;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 5 VVIVGRILSGRAIVPR 22

```

```

Db 7 VAINGKIICNPRAMIMVR 24

RESULT 12
Q84069 ID Q84069 PRELIMINARY; PRT; 29 AA.
AC Q84069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Influenza A/bellamy/42 (H1N1), non-structural protein (Seg 8), cooh
DE terminus of ns1 (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83303830; PubMed=6612993;
RA Parvin J.D., Young J.F., Palese P.;
RT "nonsense mutations affecting the lengths of the ns1 nonstructural
RT proteins of influenza A virus isolates."
RL Virology 128:512-517(1983).
DR EMBL; K00956; AAA43517.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR000256; Flu_NSI.
DR Pfam; PF00600; Flu_NSI; 1.
FT NON_TER 1
SQ SEQUENCE 29 AA; 3395 MW; A6A6A4710F3FFBEB CRC64;

Query Match 27.3%; Score 30; DB 12; Length 29;
Best Local Similarity 40.0%; Pred. No. 2.2e+03;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 14 SGRPAIVPRR 23
DB 8 NGRPPLTPKQ 17

RESULT 13
Q54460 ID Q54460 PRELIMINARY; PRT; 30 AA.
AC Q54460;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE lrp gene of Serratia marcescens (Codes for an outer membrane
DE lipoprotein) (Fragment).
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80190055; PubMed=6990409;
RA Nakamura K., Incuue M.;
RT "DNA sequence of the Serratia marcescens lipoprotein gene."
RL Proc. Natl. Acad. Sci. U.S.A. 77:1369-1373(1980).
DR EMBL; V01348; CAA24639.1; -.
DR HSP; P14178; IEO7.
DR GO; GO:0004743; F:pyruvate kinase activity; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001697; Pyruvate_kinase.
DR ProDom; PD001009; Pyruvate_kinase; 1.
FT NON_TER 1
SQ SEQUENCE 30 AA; 2965 MW; E48C7CDAACDF86C7 CRC64;

Query Match 27.3%; Score 30; DB 2; Length 30;
Best Local Similarity 41.2%; Pred. No. 2.3e+03;
Matches 7; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 KKGSVVIV--GRILSG 15

```

```
Db          4 QXGVVVVWVGALVPSG 20                               Db          8 NGRPPLTPK 16

RESULT 14
Q7VHK2
ID Q7VHK2 PRELIMINARY; PRT; 30 AA.
AC
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
DE HH0965.
GN
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Broege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Hollander R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
RL EMBL; AE017147; AAP77562.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 30 AA; 3501 MW; 8BA7EC1D7446F52B CRC64;

Query Match 27.3%; Score 30; DB 16; Length 30;
Best Local Similarity 41.7%; Pred. No. 2.3e+03;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 SVVIVGRILSG 15
DB 13 SLMLSLRYLLKG 24

RESULT 15
Q83960
ID Q83960 PRELIMINARY; PRT; 16 AA.
AC
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Influenza A/ann arbor/6/60 (H2N2), non-structural protein (Seg 8),
DE coon terminus of ns1 (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83303830; PubMed=6612993;
RA Parvin J.B., Young J.P., Palese P.;
RT "nonsense mutations affecting the lengths of the ns1 nonstructural
RT proteins of influenza A virus isolates.";
RL Virology 128:512-517(1983).
DR EMBL; K00962; AAA43516.1; -.
DR GO; GO:0003723; rRNA binding; IEA.
DR InterPro; IPR000256; Flu_Ns1.
DR Pfam; PF00600; Flu_Ns1; I.
FT NON TER 1
SQ SEQUENCE 16 AA; 1811 MW; 23F434B5AEFFAD93 CRC64;

Query Match 26.4%; Score 29; DB 12; Length 16;
Best Local Similarity 44.4%; Pred. No. 1.7e+03;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 14 SGRPAIVPR 22
:|||||:
```

Search completed: July 6, 2004, 09:43:38  
Job time : 60 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:40:55 ; Search time 19 Seconds  
(without alignments)  
62.495 Million cell updates/sec

Title: US-10-070-220-13  
Perfect score: 110  
Sequence: 1 KGSVVIVGRILSGRAIVPRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 204689

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                | Description        |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1          | 99    | 90.0        | 23     | US-08-731-336-1   | Sequence 1, Appli  |
| 2          | 99    | 90.0        | 23     | US-09-257-687-1   | Sequence 1, Appli  |
| 3          | 99    | 90.0        | 23     | US-09-881-239-8   | Sequence 8, Appli  |
| 4          | 99    | 90.0        | 23     | US-09-881-654-7   | Sequence 7, Appli  |
| 5          | 91    | 82.7        | 28     | US-09-020-846-42  | Sequence 42, Appli |
| 6          | 64    | 58.2        | 14     | US-09-011-961-6   | Sequence 6, Appli  |
| 7          | 55    | 50.0        | 14     | US-08-439-747A-14 | Sequence 14, Appli |
| 8          | 55    | 50.0        | 14     | US-08-440-409B-14 | Sequence 14, Appli |
| 9          | 55    | 50.0        | 18     | US-07-666-719-12  | Sequence 12, Appli |
| 10         | 52    | 47.3        | 12     | US-09-344-456-2   | Sequence 2, Appli  |
| 11         | 50    | 45.5        | 13     | US-08-439-747A-15 | Sequence 15, Appli |
| 12         | 50    | 45.5        | 13     | US-08-440-409B-15 | Sequence 15, Appli |
| 13         | 50    | 45.5        | 13     | US-08-853-623D-21 | Sequence 21, Appli |
| 14         | 50    | 45.5        | 13     | US-08-853-623D-30 | Sequence 30, Appli |
| 15         | 50    | 45.5        | 20     | US-08-853-623D-8  | Sequence 8, Appli  |
| 16         | 50    | 45.5        | 21     | US-08-853-623D-1  | Sequence 1, Appli  |
| 17         | 50    | 45.5        | 21     | US-08-853-623D-2  | Sequence 2, Appli  |
| 18         | 50    | 45.5        | 21     | US-08-853-623D-6  | Sequence 6, Appli  |
| 19         | 48    | 43.6        | 11     | US-09-478-479B-26 | Sequence 26, Appli |
| 20         | 48    | 43.6        | 12     | US-08-853-623D-11 | Sequence 11, Appli |
| 21         | 48    | 43.6        | 20     | US-08-617-929-1   | Sequence 1, Appli  |
| 22         | 45    | 40.9        | 10     | US-08-617-929-13  | Sequence 13, Appli |
| 23         | 45    | 40.9        | 11     | US-07-666-719-20  | Sequence 20, Appli |
| 24         | 44    | 40.0        | 20     | US-08-466-975A-9  | Sequence 9, Appli  |
| 25         | 44    | 40.0        | 20     | US-08-391-671A-9  | Sequence 9, Appli  |
| 26         | 44    | 40.0        | 20     | US-08-467-902A-9  | Sequence 9, Appli  |
| 27         | 44    | 40.0        | 20     | US-09-275-265-9   | Sequence 9, Appli  |

Query Match 90.0%; Score 99; DB 3; Length 23;  
Best Local Similarity 78.3%; Pred. No. 2.1e-09;

28 44 40.0 20 4 US-09-941-611-9 Sequence 9, Appli  
29 44 40.0 20 4 US-09-790-497A-49 Sequence 49, Appli  
30 44 40.0 20 4 US-09-790-497A-148 Sequence 148, App  
31 44 40.0 20 4 US-09-576-824A-148 Sequence 148, App  
32 44 40.0 22 2 US-08-146-028-49 Sequence 49, Appli  
33 44 40.0 22 2 US-08-146-028-148 Sequence 148, App  
34 44 40.0 22 3 US-08-723-425A-49 Sequence 49, Appli  
35 44 40.0 22 3 US-08-723-425A-148 Sequence 148, App  
36 44 40.0 22 3 US-09-112-206-49 Sequence 49, Appli  
37 44 40.0 22 3 US-09-112-206-148 Sequence 148, App  
38 44 40.0 22 3 US-09-576-824A-49 Sequence 49, Appli  
39 42 38.2 29 3 US-08-336-553A-18 Sequence 18, Appli  
40 42 38.2 29 4 US-08-439-157-18 Sequence 18, Appli  
41 42 38.2 29 4 US-09-437-895-18 Sequence 18, Appli  
42 41 37.3 9 2 US-08-146-028-262 Sequence 262, App  
43 41 37.3 9 3 US-08-723-425A-262 Sequence 262, App  
44 41 37.3 9 3 US-09-112-206-262 Sequence 262, App  
45 41 37.3 9 4 US-09-790-497A-258 Sequence 258, App

#### ALIGNMENTS

RESULT 1  
US-08-731-336-1  
; Sequence 1, Application US/08731336  
; Patent No. 615379  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Joseph L.  
; APPLICANT: Morgenstern, Kurt A.  
; APPLICANT: Lin, Chao  
; APPLICANT: Fox, Ted  
; APPLICANT: Thomson, John A.  
; TITLE OF INVENTION: CRYSTALLIZABLE COMPOSITIONS COMPRISING A  
; TITLE OF INVENTION: HEPATITIS C VIRUS NS3 PROTEASE DOMAIN/NS4A COMPLEX  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/731,336  
; FILING DATE: 18-OCT-1996  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: VP196-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; US-08-731-336-1

Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 2  
US-09-257-667-1  
; Sequence 1, Application US/09257667  
; Patent No. 6303287  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Joseph L.  
; Morgenstern, Kurt A.  
; Lin, Chao  
; Fox, Ted  
; Thomson, John A.  
; TITLE OF INVENTION: CRYSTALLIZABLE COMPOSITIONS COMPRISING A  
; HEPATITIS C VIRUS NS3 PROTEASE DOMAIN/NS4A COMPLEX  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/257,667  
; FILING DATE: 25-Feb-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/731,336  
; FILING DATE: 18-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: VPI96-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-257-667-1

Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 2.1e-09;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 3  
US-09-881-239-8  
; Sequence 8, Application US/09881239  
; Patent No. 6630298  
; GENERAL INFORMATION:

; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARCANGEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMENTO, Carlos  
; APPLICANT: COIT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY  
; FILE REFERENCE: 2302-16073 / PPI6073.003  
; CURRENT APPLICATION NUMBER: US/09/881,239  
; CURRENT FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-09-881-239-8

Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 2.1e-09;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 4  
US-09-881-654-7  
; Sequence 7, Application US/09881654  
; Patent No. 6632601  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARCANGEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMENTO, Carlos  
; APPLICANT: COIT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES  
; FILE REFERENCE: 2302-17039 / PPI7039.002  
; CURRENT APPLICATION NUMBER: US/09/881,654  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 60/212,082  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/280,811  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/280,867  
; PRIOR FILING DATE: 2001-04-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-09-881-654-7

Query Match 90.0%; Score 99; DB 4; Length 23;  
Best Local Similarity 78.3%; Pred. No. 2.1e-09;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 5  
US-09-020-846-42  
; Sequence 42, Application US/09020846  
; Patent No. 6322965

```

ADDRESS: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/011.961
FILING DATE: 23-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/IT96/00163
FILING DATE: 20-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IT RM95A000573
FILING DATE: 22-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: STEINKUHLER=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-011-961-6

Query Match 58.2%; Score 64; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00051;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gap:

QY 3 GSWIVIGRIILSGR 16
DB 1 GSWIVIGRIILSGR 14

RESULT 7
US-08-439-747A-14
Sequence 14, Application US/08439747A
Patent No. 5767233
GENERAL INFORMATION:
APPLICANT: Zhang, Rumin
APPLICANT: Murray, Michael
APPLICANT: Ramathan, Lata
TITLE OF INVENTION: Soluble, Cleavable Substrates of the Hepatitis
TITLE OF INVENTION: C Protease
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 2000 Galloping Hill Road
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.5.3
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,747A
FILING DATE: May 12, 1995
CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Paul G.  
REGISTRATION NUMBER: 32,743  
REFERENCE/DOCKET NUMBER: JB0509  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-298-5061  
TELEFAX: 908-298-5388  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
FEATURE:  
NAME/KEY: NS4A Active Mutant  
US-08-439-747A-14

Query Match 50.0%; Score 55; DB 1; Length 14;  
Best Local Similarity 78.6%; Pred. No. 0.014;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGR 16  
DB 1 GCVVIVGRIVLSGK 14

RESULT 8  
US-08-440-409B-14  
Sequence 14, Application US/08440409B  
Patent No. 5843752  
GENERAL INFORMATION:  
APPLICANT: Dasmahapatra, Bimal  
APPLICANT: Murray, Michael  
APPLICANT: Ramanathan, Lata  
APPLICANT: Butkiewicz, Nancy  
TITLE OF INVENTION: Soluble Active Hepatitis C Virus Protease  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schering Corp.  
STREET: 2000 Galloping Hill Road  
CITY: Kenilworth  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07033-0530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.5.3  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,409B  
FILING DATE: May 12, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Paul G.  
REGISTRATION NUMBER: 32,743  
REFERENCE/DOCKET NUMBER: JB0494  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-298-5388  
TELEFAX: 908-298-5388  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
FEATURE:  
NAME/KEY: NS4A Active Mutant  
US-08-440-409B-14

Query Match 50.0%; Score 55; DB 2; Length 14;

Best Local Similarity 78.6%; Pred. No. 0.014;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGR 16  
DB 1 GCVVIVGRIVLSGK 14

RESULT 9  
US-07-666-719-12  
Sequence 12, Application US/07666719  
Patent No. 5247067  
GENERAL INFORMATION:  
APPLICANT: ARIMA, Terukatsu  
APPLICANT: YAMADA, Kyoto  
APPLICANT: HATANAKA, Tadashi  
APPLICANT: NEMBA, Toshihiko  
APPLICANT: TSUJI, Masao  
TITLE OF INVENTION: PEPTIDE AND ITS USE  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: US  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/666,719  
FILING DATE: 19910422  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5247067man P.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 363-264-0X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-5940  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-666-719-12

Query Match 50.0%; Score 55; DB 1; Length 18;  
Best Local Similarity 64.3%; Pred. No. 0.019;  
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 10 RIVLSGRPAIVPR 23  
DB 1 RVVLSGKPAIIPDR 14

RESULT 10  
US-09-344-456-2  
Sequence 2, Application US/09344456A  
Patent No. 6326137  
GENERAL INFORMATION:  
APPLICANT: Hong, Zhi  
APPLICANT: Lai, Vicky C.H.  
APPLICANT: Lau, Jochson Y.N.  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE-DEPENDENT CHIMERIC  
TITLE OF INVENTION: PESTIVIRUS  
FILE REFERENCE: IN01038  
CURRENT APPLICATION NUMBER: US/09/344,456A

```
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A residues
; OTHER INFORMATION: 21-32
US-09-344-456-2

Query Match 47.3%; Score 52; DB 4; Length 12;
Best Local Similarity 91.7%; Pred. No. 0.037;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILS 14
Db 1 GSVVIVGRIVLS 12

RESULT 11
US-08-439-747A-15
; Sequence 15, Application US/08439747A
; Patent No. 5767233
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; APPLICANT: Murray, Michael
; APPLICANT: Ramanathan, Lata
; TITLE OF INVENTION: Soluble, Cleavable Substrates of the Hepatitis
; TITLE OF INVENTION: C Protease
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,747A
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: JB0509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-5061
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; FEATURE:
; NAME/KEY: NS4A Active Mutant
US-08-439-747A-15

Query Match 45.5%; Score 50; DB 1; Length 13;
Best Local Similarity 83.3%; Pred. No. 0.084;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VVIVGRILSGR 16
Db 2 VVIVGRIVLSGK 13

RESULT 12
US-08-440-409B-15
; Sequence 15, Application US/08440409B
; Patent No. 5843752
; GENERAL INFORMATION:
; APPLICANT: Dasmahapatra, Bimal
; APPLICANT: Murray, Michael
; APPLICANT: Ramanathan, Lata
; APPLICANT: Butkiewicz, Nancy
; TITLE OF INVENTION: Soluble Active Hepatitis C Virus Protease
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,409B
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: JB0494
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-5061
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; FEATURE:
; NAME/KEY: NS4A Active Mutant
US-08-440-409B-15

Query Match 45.5%; Score 50; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 0.084;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VVIVGRILSGR 16
Db 2 VVIVGRIVLSGK 13

RESULT 13
US-08-853-623D-21
; Sequence 21, Application US/08853623D
; Patent No. 5990276
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; APPLICANT: Mui, Philip
; APPLICANT: Weber, Patricia
; TITLE OF INVENTION: Synthetic Inhibitors of Hepatitis C
; TITLE OF INVENTION: Virus NS3 Protease
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
```

```
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A residues
; OTHER INFORMATION: 21-32
US-09-344-456-2

Query Match 47.3%; Score 52; DB 4; Length 12;
Best Local Similarity 91.7%; Pred. No. 0.037;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILS 14
Db 1 GSVVIVGRIVLS 12

RESULT 11
US-08-439-747A-15
; Sequence 15, Application US/08439747A
; Patent No. 5767233
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; APPLICANT: Murray, Michael
; APPLICANT: Ramanathan, Lata
; TITLE OF INVENTION: Soluble, Cleavable Substrates of the Hepatitis
; TITLE OF INVENTION: C Protease
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,747A
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: JB0509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-5061
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; FEATURE:
; NAME/KEY: NS4A Active Mutant
US-08-439-747A-15

Query Match 45.5%; Score 50; DB 1; Length 13;
Best Local Similarity 83.3%; Pred. No. 0.084;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VVIVGRILSGR 16
Db 2 VVIVGRIVLSGK 13

RESULT 12
US-08-440-409B-15
; Sequence 15, Application US/08440409B
; Patent No. 5843752
; GENERAL INFORMATION:
; APPLICANT: Dasmahapatra, Bimal
; APPLICANT: Murray, Michael
; APPLICANT: Ramanathan, Lata
; APPLICANT: Butkiewicz, Nancy
; TITLE OF INVENTION: Soluble Active Hepatitis C Virus Protease
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,409B
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: JB0494
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-5061
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; FEATURE:
; NAME/KEY: NS4A Active Mutant
US-08-440-409B-15

Query Match 45.5%; Score 50; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 0.084;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VVIVGRILSGR 16
Db 2 VVIVGRIVLSGK 13

RESULT 13
US-08-853-623D-21
; Sequence 21, Application US/08853623D
; Patent No. 5990276
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; APPLICANT: Mui, Philip
; APPLICANT: Weber, Patricia
; TITLE OF INVENTION: Synthetic Inhibitors of Hepatitis C
; TITLE OF INVENTION: Virus NS3 Protease
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
```



; OTHER INFORMATION: shown), preferably the amino acid residues at positions  
; OTHER INFORMATION: 8 - 20 are D- amino acid residues.  
US-08-853-623D-8

Query Match 45.5%; Score 50; DB 2; Length 20;  
Best Local Similarity 83.3%; Pred. No. 0.14; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 WVVGRIILSGR 16  
||| ||| :|||  
Db 9 WVVGRIIVLSGX 20

Search completed: July 6, 2004, 09:44:36  
Job time : 20 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:43:44 ; Search time 62 Seconds  
(without alignments)  
115.476 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110  
Sequence: 1 KGSVVIVGRILSGRPAIVPRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311293816 residues

Total number of hits satisfying chosen parameters: 284211

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description       |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1          | 99    | 90.0        | 23     | 9 US-09-881-654-7    | Sequence 7, Appli |
| 2          | 99    | 90.0        | 23     | 9 US-09-881-239-8    | Sequence 8, Appli |
| 3          | 99    | 90.0        | 23     | 12 US-10-637-323-7   | Sequence 7, Appli |
| 4          | 99    | 90.0        | 23     | 12 US-09-728-653-6   | Sequence 6, Appli |
| 5          | 99    | 90.0        | 23     | 14 US-10-010-184A-7  | Sequence 7, Appli |
| 6          | 99    | 90.0        | 23     | 14 US-10-015-328-6   | Sequence 6, Appli |
| 7          | 99    | 90.0        | 23     | 16 US-10-643-853-8   | Sequence 8, Appli |
| 8          | 78    | 70.9        | 30     | 12 US-10-296-734-632 | Sequence 632, App |
| 9          | 74    | 67.3        | 17     | 12 US-10-221-943-31  | Sequence 31, Appl |
| 10         | 70    | 63.6        | 16     | 14 US-10-319-402-19  | Sequence 19, Appl |
| 11         | 66    | 60.0        | 30     | 12 US-10-296-734-630 | Sequence 630, App |
| 12         | 60    | 54.5        | 14     | 14 US-10-319-402-3   | Sequence 3, Appli |
| 13         | 50    | 45.5        | 12     | 14 US-10-319-402-21  | Sequence 21, Appl |
| 14         | 48    | 43.6        | 11     | 9 US-09-965-594-26   | Sequence 26, Appl |
| 15         | 48    | 43.6        | 12     | 14 US-10-319-402-2   | Sequence 2, Appli |

|    |    |      |    |    |                    |                    |
|----|----|------|----|----|--------------------|--------------------|
| 16 | 46 | 41.8 | 11 | 14 | US-10-319-402-37   | Sequence 37, Appli |
| 17 | 44 | 40.0 | 20 | 9  | US-09-941-611-9    | Sequence 9, Appli  |
| 18 | 44 | 40.0 | 20 | 10 | US-09-899-046-102  | Sequence 102, App  |
| 19 | 44 | 40.0 | 20 | 10 | US-09-898-046-103  | Sequence 103, App  |
| 20 | 44 | 40.0 | 20 | 10 | US-09-878-281-102  | Sequence 102, App  |
| 21 | 44 | 40.0 | 20 | 10 | US-09-878-281-103  | Sequence 103, App  |
| 22 | 44 | 40.0 | 20 | 12 | US-09-873-224-102  | Sequence 102, App  |
| 23 | 44 | 40.0 | 20 | 12 | US-09-873-224-103  | Sequence 103, App  |
| 24 | 44 | 40.0 | 20 | 14 | US-10-044-995-9    | Sequence 9, Appli  |
| 25 | 43 | 39.1 | 11 | 14 | US-10-319-402-22   | Sequence 22, Appl  |
| 26 | 42 | 38.2 | 10 | 14 | US-10-319-402-35   | Sequence 35, Appl  |
| 27 | 40 | 36.4 | 18 | 14 | US-10-371-540-1    | Sequence 1, Appli  |
| 28 | 40 | 36.4 | 20 | 14 | US-10-397-411-15   | Sequence 15, Appl  |
| 29 | 39 | 35.5 | 10 | 14 | US-10-319-402-36   | Sequence 36, Appl  |
| 30 | 38 | 34.5 | 9  | 14 | US-10-319-402-16   | Sequence 16, Appl  |
| 31 | 38 | 34.5 | 9  | 14 | US-10-319-402-33   | Sequence 33, Appl  |
| 32 | 38 | 34.5 | 20 | 10 | US-09-899-046-97   | Sequence 97, Appl  |
| 33 | 38 | 34.5 | 20 | 10 | US-09-878-281-97   | Sequence 97, Appl  |
| 34 | 38 | 34.5 | 20 | 12 | US-09-873-224-97   | Sequence 97, Appl  |
| 35 | 37 | 33.6 | 30 | 9  | US-09-925-299-1432 | Sequence 1432, Ap  |
| 36 | 37 | 33.6 | 30 | 10 | US-09-925-299-1432 | Sequence 1432, Ap  |
| 37 | 36 | 32.7 | 9  | 14 | US-10-319-402-34   | Sequence 34, Appl  |
| 38 | 36 | 32.7 | 20 | 10 | US-09-899-046-98   | Sequence 98, Appl  |
| 39 | 35 | 32.7 | 20 | 10 | US-09-878-281-98   | Sequence 98, Appl  |
| 40 | 35 | 32.7 | 20 | 12 | US-09-873-224-98   | Sequence 98, Appl  |
| 41 | 35 | 31.8 | 19 | 14 | US-10-161-791-185  | Sequence 185, App  |
| 42 | 35 | 31.8 | 23 | 12 | US-09-747-287-178  | Sequence 178, App  |
| 43 | 35 | 31.8 | 23 | 12 | US-09-874-350A-144 | Sequence 144, App  |
| 44 | 35 | 31.8 | 26 | 12 | US-10-633-983-114  | Sequence 114, App  |
| 45 | 35 | 31.8 | 26 | 16 | US-10-443-622-73   | Sequence 73, Appl  |

#### ALIGNMENTS

##### RESULT 1

US-09-881-654-7  
; Sequence 7, Application US/09881654  
; Patent No. US2002014685A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARCANGEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMENTO, Carlos  
; APPLICANT: COLT Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES  
; FILE REFERENCE: 2302-17039 / PP17039.002  
; CURRENT APPLICATION NUMBER: US/09/881,654  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 60/212,082  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/280,811  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/280,867  
; PRIOR FILING DATE: 2001-04-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-09-881-654-7

Query Match 90.0%; Score 99; DB 9; Length 23;  
Best local similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KGSVVIVGRILSGRPAIVPRR 23  
|||||:|||||:|||||:|||||:  
Db 1 KGSVVIVGRILSGRPAIVPRR 23

```

US-09-881-239-8
RESULT 2
; Sequence 8, Application US/09881239
; Publication No. US20020192639A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
; FILE REFERENCE: 2302-16073 / PPI16073.003
; CURRENT APPLICATION NUMBER: US/09/881,239
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide
US-09-881-239-8
Query Match 90.0%; Score 99; DB 9; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23

US-10-637-323-7
RESULT 3
; Sequence 7, Application US/10637323
; Publication No. US20040063092A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PPI17039.002
; CURRENT APPLICATION NUMBER: US/10/637,323
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: US/09/881,654
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/280,867
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide
US-10-637-323-7
Query Match 90.0%; Score 99; DB 12; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23

US-09-881-239-8
RESULT 2
; Sequence 8, Application US/09881239
; Publication No. US20020192639A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
; FILE REFERENCE: 2302-16073 / PPI16073.003
; CURRENT APPLICATION NUMBER: US/09/881,239
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide
US-09-881-239-8
Query Match 90.0%; Score 99; DB 9; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23

US-10-010-184A-7
RESULT 5
; Sequence 7, Application US/10010184A
; Publication No. US20030008928A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Pharma Company
; APPLICANT: Priestly, et al.
; TITLE OF INVENTION: No. US20030008928A1 Lactam Inhibitors of Hepatitis C Virus NS3
; FILE REFERENCE: PH-7087-A
; CURRENT APPLICATION NUMBER: US/10/010,184A
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 09/626,286
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A pept
; OTHER INFORMATION: ide synthesizer using readily available materials well known to o
; OTHER INFORMATION: rdinarily skilled artisans
US-10-010-184A-7
Query Match 90.0%; Score 99; DB 14; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23

US-09-728-653-6
RESULT 4
; Sequence 6, Application US/09728653
; Publication No. US20020123468A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Han, Amy Qi
; APPLICANT: Glunz, Peter W.
; TITLE OF INVENTION: Alpha-Ketoamide Inhibitors of Hepatitis C Virus NS3 Protease
; FILE REFERENCE: PH-7118
; CURRENT APPLICATION NUMBER: US/09/728,653
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/168,998
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A pept
; OTHER INFORMATION: ide synthesizer using readily available materials well known to o
; OTHER INFORMATION: rdinarily skilled artisans
US-09-728-653-6
Query Match 90.0%; Score 99; DB 12; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23

US-10-010-184A-7
RESULT 5
; Sequence 7, Application US/10010184A
; Publication No. US20030008928A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Pharma Company
; APPLICANT: Priestly, et al.
; TITLE OF INVENTION: No. US20030008928A1 Lactam Inhibitors of Hepatitis C Virus NS3
; FILE REFERENCE: PH-7087-A
; CURRENT APPLICATION NUMBER: US/10/010,184A
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 09/626,286
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A pept
; OTHER INFORMATION: ide synthesizer using readily available materials well known to o
; OTHER INFORMATION: rdinarily skilled artisans
US-10-010-184A-7
Query Match 90.0%; Score 99; DB 14; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPR 23
Db 1 KKGSVVIVGRIVLSGRPAIIPK 23
```

RESULT 6  
US-10-015-328-6  
; Sequence 6, Application US/10015328  
; Publication No. US20030100768A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; APPLICANT: Han, Amy Q  
; APPLICANT: Glunz, Peter W.  
; TITLE OF INVENTION: Imidazolidinones and Their Related Derivatives as Hepatitis C Virus Inhibitors  
; FILE REFERENCE: PH-7203  
; CURRENT APPLICATION NUMBER: US/10/015,328  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255,168  
; PRIOR FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A peptide synthesizer using readily available materials well known to those skilled in the art.  
; OTHER INFORMATION: rdinarily skilled artisans  
US-10-015-328-6

Query Match 90.0%; Score 99; DB 14; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 7  
US-10-643-853-8  
; Sequence 8, Application US/10643853  
; Publication No. US2004009622A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARANGEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMTO, Carlos  
; APPLICANT: COIT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY  
; FILE REFERENCE: 2302-16073 / PP16073.003  
; CURRENT APPLICATION NUMBER: US/10/643,853  
; CURRENT FILING DATE: 2003-08-19  
; PRIOR APPLICATION NUMBER: US/09/881,239  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-10-643-853-8

Query Match 90.0%; Score 99; DB 16; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGRPAIVPRR 23  
Db 1 KKGSVVIVGRILSGRPAIVPRR 23

RESULT 8  
US-10-296-734-632  
; Sequence 632, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A  
; TITLE OF INVENTION: Synthetic molecules and uses therefor  
; FILE REFERENCE: Savine  
; CURRENT APPLICATION NUMBER: US/10/296,734  
; CURRENT FILING DATE: 2003-08-04  
; PRIOR APPLICATION NUMBER: AU PQ7761/00  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 1507  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 632  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: HepC la segment 114  
US-10-296-734-632

Query Match 70.9%; Score 78; DB 12; Length 30;  
Best Local Similarity 78.9%; Pred. No. 4.3e-05;  
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 5 WVIVGRILSGRPAIVPRR 23  
Db 2 WVIVGRILSGRPAIVPRR 20

RESULT 9  
US-10-221-943-31  
; Sequence 31, Application US/10221943  
; Publication No. US20040054134A1  
; GENERAL INFORMATION:  
; APPLICANT: Steinkuhler, Christian  
; APPLICANT: Pallao, Michele  
; APPLICANT: Lahm, Armin  
; TITLE OF INVENTION: HCV NS2/3 FRAGMENTS AND USES THEREOF  
; FILE REFERENCE: IT0031P  
; CURRENT APPLICATION NUMBER: US/10/221,943  
; CURRENT FILING DATE: 2002-09-17  
; PRIOR APPLICATION NUMBER: PCT/IB01/0052  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: GB 0006537.5  
; PRIOR FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Pep4AK  
; FEATURE:  
; NAME/KEY: AMIDATION  
; LOCATION: (17)...(17)  
US-10-221-943-31

Query Match 67.3%; Score 74; DB 12; Length 17;  
Best Local Similarity 100.0%; Pred. No. 9.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVVIVGRILSGR 16  
Db 2 KKGSVVIVGRILSGR 17

RESULT 10  
US-10-319-402-19  
; Sequence 19, Application US/10319402

```
/ Publication No. US20030176689A1
/ GENERAL INFORMATION:
/ APPLICANT: Joyce, Michael
/ APPLICANT: Williams, Mark
/ APPLICANT: Hinds Gaul, Ole
/ TITLE OF INVENTION: Inhibitors of Hepatitis C Virus Protease
/ FILE REFERENCE: UALB-002
/ CURRENT APPLICATION NUMBER: US/10/319,402
/ PRIOR FILING DATE: 2002-12-13
/ PRIOR APPLICATION NUMBER: US 60/340,574
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 19
/ LENGTH: 16
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: binding and activation domain
/ NAME/KEY: PEPTIDE
/ LOCATION: (1)..(16)
/ OTHER INFORMATION: This sequence is shown in the specification using one-letter code
US-10-319-402-19

Query Match      63.6%; Score 70; DB 14; Length 16;
Best Local Similarity 87.5%; Pred. No. 0.00039;
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KGSVVIVGRILSGR 16
Db 1 KGSVVIVGRIVLSGK 16

RESULT 11
US-10-296-734-630
/ Sequence 630, Application US/10296734
/ Publication No. US20040054137A1
/ GENERAL INFORMATION:
/ APPLICANT: Thompson, Scott A
/ APPLICANT: Ramshaw, Ian A
/ TITLE OF INVENTION: Synthetic molecules and uses therefor
/ FILE REFERENCE: Savine
/ CURRENT APPLICATION NUMBER: US/10/296,734
/ PRIOR FILING DATE: 2003-08-04
/ PRIOR APPLICATION NUMBER: AU PQ7761/00
/ PRIOR FILING DATE: 2000-05-26
/ NUMBER OF SEQ ID NOS: 1507
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 630
/ LENGTH: 30
/ TYPE: PRT
/ ORGANISM: Artificial
/ FEATURE:
/ OTHER INFORMATION: HepC 1a segment 112
US-10-296-734-630

Query Match      60.0%; Score 66; DB 12; Length 30;
Best Local Similarity 81.2%; Pred. No. 0.0033;
Matches 13; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGRPA 18
Db 15 GCVVIVGRIVLSGKPA 30

RESULT 12
US-10-319-402-3
/ Sequence 3, Application US/10319402
/ Publication No. US20030176689A1
/ GENERAL INFORMATION:
/ APPLICANT: Joyce, Michael
/ APPLICANT: Williams, Mark
/ APPLICANT: Hinds Gaul, Ole
/ TITLE OF INVENTION: Inhibitors of Hepatitis C Virus Protease
/ FILE REFERENCE: UALB-002
/ CURRENT APPLICATION NUMBER: US/10/319,402
/ PRIOR FILING DATE: 2002-12-13
/ PRIOR APPLICATION NUMBER: US 60/340,574
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 14
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Native amino acid sequence of NS4A's binding and activation domain
/ NAME/KEY: PEPTIDE
/ LOCATION: (1)..(14)
/ OTHER INFORMATION: This sequence is shown in the specification using one-letter code
US-10-319-402-3

Query Match      54.5%; Score 60; DB 14; Length 14;
Best Local Similarity 85.7%; Pred. No. 0.013;
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGR 16
Db 1 GSVVIVGRIVLSGK 14

RESULT 13
US-10-319-402-21
/ Sequence 21, Application US/10319402
/ Publication No. US20030176689A1
/ GENERAL INFORMATION:
/ APPLICANT: Joyce, Michael
/ APPLICANT: Williams, Mark
/ APPLICANT: Hinds Gaul, Ole
/ TITLE OF INVENTION: Inhibitors of Hepatitis C Virus Protease
/ FILE REFERENCE: UALB-002
/ CURRENT APPLICATION NUMBER: US/10/319,402
/ CURRENT FILING DATE: 2002-12-13
/ PRIOR APPLICATION NUMBER: US 60/340,574
/ PRIOR FILING DATE: 2001-12-14
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 21
/ LENGTH: 12
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: binding and activation domain
/ NAME/KEY: PEPTIDE
/ LOCATION: (1)..(12)
/ OTHER INFORMATION: This sequence is shown in the specification using one-letter code
US-10-319-402-21

Query Match      45.5%; Score 50; DB 14; Length 12;
Best Local Similarity 83.3%; Pred. No. 0.4;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VVIVGRILSGR 16
Db 1 VVIVGRIVLSGK 12

RESULT 14
US-09-965-594-26
/ Sequence 26, Application US/09965594
```

```
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekkind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yagun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-965-594-26

Query Match 43.6%; Score 48; DB 9; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.75;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GSVVIVGRILL 13
Db 1 GSVVIVGRIVL 11

RESULT 15
US-10-319-402-2
; Sequence 2, Application US/10319402
; Publication No. US20030176689A1
; GENERAL INFORMATION:
; APPLICANT: Joyce, Michael
; APPLICANT: Williams, Mark
; APPLICANT: Hindsgaul, Ole
; APPLICANT: Tyrell, D. Lorne
; TITLE OF INVENTION: Inhibitors of Hepatitis C Virus Protease
; FILE REFERENCE: UALB-002
; CURRENT APPLICATION NUMBER: US/10/319,402
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 60/340,574
; PRIOR FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Binding and activation domain
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(12)
; OTHER INFORMATION: This sequence is shown in the specification using one-letter code
US-10-319-402-2

Query Match 43.6%; Score 48; DB 14; Length 12;
Best Local Similarity 90.9%; Pred. No. 0.82;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 WVIVGRIVLSG 15
Db 2 WVIVGRIVLSG 12
```

Search completed: July 6, 2004, 09:49:50  
Job time : 62 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:28:15 ; Search time 37 Seconds  
(without alignments)  
175.638 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110

Sequence: 1 KKGSVIVGRILLGRPAIVPRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 110   | 100.0       | 23     | 4     | AAB96855 Hepatitis |
| 2          | 103   | 93.6        | 32     | 4     | AAB96856 Hepatitis |
| 3          | 99    | 90.0        | 23     | 2     | AAW40552 Synthetic |
| 4          | 99    | 90.0        | 23     | 4     | AAW52536 Peptide K |
| 5          | 99    | 90.0        | 23     | 4     | AAB67391 KNS4a pe  |
| 6          | 99    | 90.0        | 23     | 4     | AAB66371 Hepatitis |
| 7          | 99    | 90.0        | 23     | 4     | AAW64158 KNS4a pe  |
| 8          | 99    | 90.0        | 23     | 5     | ABG31914 KNS4a pe  |
| 9          | 99    | 90.0        | 23     | 5     | AAU76376 Hepatitis |
| 10         | 99    | 90.0        | 23     | 5     | AAE18687 NS4A pept |
| 11         | 99    | 90.0        | 23     | 5     | ABG32508 Peptide K |
| 12         | 99    | 90.0        | 23     | 6     | ABG72264 Hepatitis |
| 13         | 99    | 90.0        | 23     | 7     | ABU61521 Hepatitis |
| 14         | 99    | 90.0        | 23     | 7     | ADC06772 HCV NS4A  |
| 15         | 93    | 84.5        | 34     | 2     | AAW82856 NS3 serin |
| 16         | 93    | 84.5        | 54     | 2     | AAW82855 NS4A prot |
| 17         | 93    | 84.5        | 54     | 2     | AAW37808 Nonstruct |
| 18         | 93    | 84.5        | 54     | 2     | AAW17898 Native HC |
| 19         | 93    | 84.5        | 86     | 2     | AAW09051 Hepatitis |
| 20         | 93    | 84.5        | 87     | 2     | AAW95545 HCV II ch |
| 21         | 93    | 84.5        | 373    | 2     | AAW41436 PT-NANBH  |
| 22         | 93    | 84.5        | 1250   | 2     | AAW12599 Portion o |
| 23         | 93    | 84.5        | 1985   | 6     | ABU09574 HCV Met-N |
| 24         | 93    | 84.5        | 1985   | 6     | ABU09575 HCV Met-N |
| 25         | 93    | 84.5        | 2201   | 2     | AAW01680 HCV NS2-N |

|    |    |      |      |   |          |                    |
|----|----|------|------|---|----------|--------------------|
| 26 | 93 | 84.5 | 2307 | 3 | AAW70064 | Aay70064 Recombina |
| 27 | 93 | 84.5 | 2307 | 3 | AAW70065 | Aay70065 Recombina |
| 28 | 93 | 84.5 | 2307 | 3 | AAW70066 | Aay70066 Recombina |
| 29 | 93 | 84.5 | 2354 | 2 | AAW41435 | PT-NANBH           |
| 30 | 93 | 84.5 | 3010 | 2 | AAW20111 | Non-A, no          |
| 31 | 93 | 84.5 | 3010 | 2 | AAW20091 | Non-A, no          |
| 32 | 93 | 84.5 | 3010 | 2 | AAW06423 | Non-A, no          |
| 33 | 92 | 83.6 | 320  | 2 | AAW29907 | HCV antiq          |
| 34 | 92 | 83.6 | 320  | 2 | AAW29908 | HCV antiq          |
| 35 | 92 | 83.6 | 916  | 2 | AAW82693 | HCV parti          |
| 36 | 92 | 83.6 | 923  | 4 | AAW82696 | HCV parti          |
| 37 | 92 | 83.6 | 1736 | 4 | AAW36932 | Hepatitis          |
| 38 | 92 | 83.6 | 2201 | 5 | ABG30586 | Hepatitis          |
| 39 | 92 | 83.6 | 2201 | 5 | ABG30583 | Hepatitis          |
| 40 | 92 | 83.6 | 2201 | 5 | ABG30588 | Hepatitis          |
| 41 | 92 | 83.6 | 3010 | 2 | AAW34580 | Human hep          |
| 42 | 92 | 83.6 | 3010 | 2 | AAW53417 | Blood tra          |
| 43 | 92 | 83.6 | 3011 | 2 | AAW34468 | Encoded b          |
| 44 | 92 | 83.6 | 3014 | 2 | AAW35207 | Hepatitis          |
| 45 | 92 | 83.6 | 3014 | 2 | AAW54099 | NANBHV E1          |

ALIGNMENTS

RESULT 1  
AAB96855  
ID AAB96855 standard; peptide; 23 AA.  
XX  
AC AAB96855;  
XX  
DT 06-JUL-2001 (first entry)  
XX  
DE Hepatitis C virus NS2/3 cleavage inhibitory peptide SEQ ID NO: 13.  
XX  
KW Hepatitis C virus; HCV; NS2; NS3; inhibitory peptide; cleavage;  
KW replication inhibition; chimpanzee; human; infection; gene therapy.  
XX  
OS Hepatitis C virus.  
XX  
PH Key Location/Qualifiers  
FT Modified-site 23  
FT /label= OTHER  
FT /note= "C-terminal amide"  
XX  
PN WO200116379-A1.  
XX  
PD 08-MAR-2001.  
XX  
PF 25-AUG-2000; 2000WO-US023444.  
XX  
PR 30-AUG-1999; 99US-0151395P.  
XX  
PA (MERI ) MERCK & CO INC.  
XX  
PI Darke PL, Jacobs AR, Kuo LC;  
XX  
DR WPI; 2001-343059/36.  
XX  
PT Inhibiting hepatitis C virus (HCV) replication in HCV infected cell, or  
PT in a patient or treating a patient for HCV infection comprises inhibiting  
PT autocleavage of NS2/3.  
XX  
PS Claim 29; Page 34; 50pp; English.  
XX  
CC The present invention describes methods and compositions capable of  
CC preventing the replication of hepatitis C virus (HCV), involving  
CC administering a compound which inhibits NS2/3 autocleavage. Also provided  
CC are peptides capable of inhibiting this cleavage step, of which this  
CC sequence is an example. These are useful in the treatment of HCV  
CC infection in humans and chimpanzees, and in research applications, for  
CC example in studying the stabilisation of NS2/3, the effects of NS2/3 on  
CC HCV polyprotein processing and the effects of inhibiting NS2/3

```

CC autocleavage
XX
SQ Sequence 23 AA;
  Query Match      100.0%; Score 110; DB 4; Length 23;
  Best Local Similarity 100.0%; Pred. No. 2.9e-10;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVIVIGRIILSGRPAIVPRR 23
DB 1 KKGSVIVIGRIILSGRPAIVPRR 23

RESULT 2
AAB96856
ID AAB96856 standard; peptide; 32 AA.
XX
AC AAB96856;
XX
DT 06-JUL-2001 (first entry)
XX
DE Hepatitis C virus NS2/3 cleavage inhibitory peptide SEQ ID NO: 14.
XX
KW Hepatitis C virus; HCV; NS2; NS3; inhibitory peptide; cleavage;
KW replication inhibition; chimpanzee; human; infection; gene therapy.
XX
OS Hepatitis C virus.
XX
FN WO200116379-A1.
XX
PD 08-MAR-2001.
XX
PF 25-AUG-2000; 2000WO-US023444.
XX
PR 30-AUG-1999; 99US-0151395P.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Darke PL, Jacobs AR, Kuo LC;
XX
DR WPI; 2001-343059/36.
XX
PT Inhibiting hepatitis C virus (HCV) replication in HCV infected cell, or
PT in a patient or treating a patient for HCV infection comprises inhibiting
PT autocleavage of NS2/3.
XX
PS Claim 29; Page 34; 50pp; English.
XX
CC The present invention describes methods and compositions capable of
CC preventing the replication of hepatitis C virus (HCV), involving
CC administering a compound which inhibits NS2/3 autocleavage. Also provided
CC are peptides capable of inhibiting this cleavage step, of which this
CC sequence is an example. These are useful in the treatment of HCV
CC infection in humans and chimpanzees, and in research applications, for
CC example in studying the stabilisation of NS2/3, the effects of NS2/3 on
CC HCV polyprotein processing and the effects of inhibiting NS2/3
CC autocleavage
XX
SQ Sequence 32 AA;
  Query Match      93.6%; Score 103; DB 4; Length 32;
  Best Local Similarity 95.7%; Pred. No. 5.3e-09;
  Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKGSVIVIGRIILSGRPAIVPRR 23
DB 1 KKGSVIVIGRIILSGRPAIVPRR 23

RESULT 3
AAW40552
ID AAW40552 standard; peptide; 23 AA.
XX

```

---

```

AC AAW40552;
XX
DT 20-NOV-1998 (first entry)
XX
DE Synthetic nonstructural peptide SNS4A.
XX
KW Synthetic Hepatitis C nonstructural protein; SNS4A peptide; cofactor;
KW NS3 protease.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
FN WO9811134-A1.
XX
PD 19-MAR-1998.
XX
PF 12-SEP-1997; 97WO-US016182.
XX
PR 12-SEP-1996; 96US-0025274P.
PR 18-OCT-1996; 96US-00731336.
XX
PA (VERT-) VERTEX PHARM INC.
XX
PI Kim JL, Morgenstern KA, Lin C, Fox T, Thomson JA;
XX
DR WPI; 1998-250953/22.
XX
PT New hepatitis C virus crystal compositions - comprising a HCV NS3-like
PT polypeptide complexed with a NS4A-like polypeptide, used particularly for
PT drug design.
XX
PS Claim 4; Page 30; 97pp; English.
XX
CC This is the amino acid sequence of the novel SNS4A (synthetic Hepatitis C
CC nonstructural protein 4A) peptide. It acts as a cofactor for the NS3
CC protease in order to achieve proteolytic processing of Hepatitis C virus
CC (HCV) nonstructural proteins. It is used in the method of the invention
CC as part of a device which can be used to provide information for the
CC design of drugs for the treatment of HCV infection. They can also be used
CC for determining the 3-dimensional structure of molecules or molecular
CC complexes which contain at least some structurally similar features to a
CC HCV NS3 serine protease domain
XX
SQ Sequence 23 AA;
  Query Match      90.0%; Score 99; DB 2; Length 23;
  Best Local Similarity 78.3%; Pred. No. 1.6e-08;
  Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVIVIGRIILSGRPAIVPRR 23
DB 1 KKGSVIVIGRIILSGRPAIVPRR 23

RESULT 4
AAM52536
ID AAM52536 standard; peptide; 23 AA.
XX
AC AAM52536;
XX
DT 31-JAN-2002 (first entry)
XX
DE Peptide KKSNS4a used in an enzyme assay.
XX
KW Virucide; pyrrolopyrazinone derivative; Hepatitis C virus inhibitor;
KW nonstructural 3 protease; NS3 protease; viral infection.
XX
OS Synthetic.
XX
FN WO200164678-A2.
XX
PD 07-SEP-2001.
XX

```



```
PF 28-FEB-2001; 2001WO-US006269.
XX
PR 29-FEB-2000; 2000US-0185618P.
XX
PA (DUPO ) DUPONT PHARM CO.
XX
PI Zhang X, Han W;
XX
DR WPI; 2001-656752/75.
XX
XX New pyrrolopyrazinone derivatives useful for treating Hepatitis C virus
PT infection are NS3 protease inhibitors.
PT
PS Disclosure; Page 130; 191pp; English.
XX
XX The present invention relates to a novel pyrrolopyrazinone derivative,
CC its stereoisomer or salt. It was found that the derivative is a Hepatitis
CC C virus (HCV) nonstructural (NS) 3 protease inhibitor. The derivative is
CC useful for the manufacture of a medicament for the treatment of HCV and
CC in therapy for treating HCV infection. The present peptide was used to
CC illustrate the present invention
XX
XX Sequence 23 AA;
SQ

Query Match 90.0%; Score 99; DB 4; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23
   |||||:||||:||||:||||:
DB 1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 5
AAB67391
ID AAB67391 standard; peptide; 23 AA.
XX
XX AAB67391;
AC
XX 09-APR-2001 (first entry)
DT Hepatitis C virus protease inhibitor related peptide #1.
CC
DE Hepatitis C virus; protease; boronic acid; inhibitor; liver cirrhosis;
XX
KW liver cancer; NS3; antiviral agent.
XX
XX Unidentified.
OS
XX WO200102424-A2.
PN
XX 11-JAN-2001.
PD
XX 07-JUL-2000; 2000WO-US018655.
PF
XX 07-JUL-1999; 99US-0142561P.
PR
XX (DUPO ) DU PONT PHARM CO.
PA
XX Kettner CA, Jagannathan S, Forsyth TP;
PI
XX WPI; 2001-103001/11.
DR
XX
XX New boronic acid derivatives, optionally containing peptides, used to
PT treat hepatitis C infections, are hepatitis C viral protease inhibitors.
XX
XX Example 60; Page 208; 258pp; English.
PS
XX
XX The present invention provides a number of boronic acid derivatives which
CC act as inhibitors of the hepatitis C virus NS3 protease enzyme. They can
CC be used to treat infection by the virus, which can cause liver cirrhosis
CC and liver cancer
CC
XX Sequence 23 AA;
SQ

Query Match 90.0%; Score 99; DB 4; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23
   |||||:||||:||||:||||:
DB 1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 7
AAG64158
ID AAG64158 standard; peptide; 23 AA.
XX
XX AAG64158;
AC
XX
XX 19-OCT-2001 (first entry)
DT
XX
XX KKS4a peptide.
DE
XX Hepatitis C virus; HCV; NS3 protease; alpha-ketoamide inhibitor;
KW virucide; hepatotropic; antiinflammatory; viral infection; KKS4a.
XX
```

OS Synthetic.  
 XX WO200140262-A1.  
 PN 07-JUN-2001.  
 PD 01-DEC-2000; 2000WO-US032677.  
 XX 03-DEC-1999; 99US-0168998P.  
 PR (DUPO ) DU PONT PHARM CO.  
 XX Han W;  
 PI WPI; 2001-464936/50.  
 XX New ketoamide derivatives useful for treating infections e.g. hepatitis C virus.  
 PT Disclosure; Page 195; 282pp; English.  
 XX The invention relates to novel ketoamide and ketoester derivatives for use as inhibitors of hepatitis C virus (HCV) NS3 protease inhibitors. The compounds are useful for treating viral infections e.g. hepatitis C virus. The present sequence was used in an experiment measuring the effect of an inhibitor on the rate of hydrolysis of an ester substrate

XX Query Match 90.0%; Score 99; DB 4; Length 23;  
 XX Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 XX Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 XX Sequence 23 AA;  
 XX 1 KKGSVVIVGRILSGSPAIVPR 23  
 XX 1 KKGSVVIVGRILSGKPAIIPKX 23

RESULT 8  
 ABG31914  
 ID ABG31914 standard; peptide; 23 AA.  
 XX AC ABG31914;  
 XX 05-NOV-2002 (first entry)  
 XX KONS4a peptide.  
 XX HCV; hepatitis C; imidazolidinone; serine protease inhibitor; virucide; hepatotropic; antiinflammatory; NS3 protease; KONS4a; growth inhibitor; viral infection; blood plasma processing.  
 XX Synthetic.  
 XX WO200248157-A2.  
 XX 20-JUN-2002.  
 XX 12-DEC-2001; 2001WO-US047916.  
 XX 13-DEC-2000; 2000US-0255168P.  
 XX (BRIM ) BRISTOL MYERS SQUIBB PHARMA CO.  
 XX Han Q;  
 XX WPI; 2002-599498/64.  
 XX New imidazolidinones useful as serine protease inhibitors in the treatment of e.g. viral infection.  
 XX Example 20; Page 112; 173pp; English.

CC This invention relates to novel imidazolidinones or their stereoisomers, salts or prodrugs which are useful as serine protease inhibitors. The imidazolidinones of the invention may have virucide, hepatotropic, or antiinflammatory activities and may be used as a serine protease inhibitor (preferably Hepatitis C virus (HCV) NS3 protease inhibitor) or a HCV growth inhibitor. Compounds of the invention are useful for treating viral infection e.g. hepatitis C virus (HCV) infection and as a reagent used as inhibitors of HCV protease in the processing of blood plasma for diagnostic and other commercial purposes. The imidazolidinones of the invention inhibit HCV NS3 protease and/or HCV growth and thus can be used in the blood plasma assay. The present sequence represents the KONS4a peptide used in enzyme assay experiments in the examples of the specification

XX Query Match 90.0%; Score 99; DB 5; Length 23;  
 XX Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 XX Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 XX Sequence 23 AA;  
 XX 1 KKGSVVIVGRILSGSPAIVPR 23  
 XX 1 KKGSVVIVGRILSGKPAIIPKX 23

RESULT 9  
 AAU76376  
 ID AAU76376 standard; peptide; 23 AA.  
 XX AC AAU76376;  
 XX 08-MAY-2002 (first entry)  
 XX Hepatitis C virus (non-structural protein) NS4A peptide sequence.  
 XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion; immunoassay solid support; multiple epitope fusion antigen; MEFA; non-structural protein; NS4A.  
 XX Hepatitis C virus.  
 XX WO200196870-A2.  
 XX 20-DEC-2001.  
 XX 14-JUN-2001; 2001WO-US019156.  
 XX 15-JUN-2000; 2000US-0212082P.  
 XX 02-APR-2001; 2001US-0280811P.  
 XX 02-APR-2001; 2001US-0280867P.  
 XX (CHIR ) CHIRON CORP.  
 XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D; Medina-Selby A;  
 XX WPI; 2002-090228/12.  
 XX Immunoassay solid support, useful for detecting hepatitis C virus infection in biological sample, comprises HCV NS3/4a conformational epitope and multiple epitope fusion antigen bound to the support.  
 XX Example 3; Page 48; 92pp; English.  
 XX The present invention relates to a new immunoassay solid support consisting essentially of at least one hepatitis C virus (HCV) NS3/4a conformational epitope and a multiple epitope fusion antigen (MEFA), bound to the support. The NS3/4a conformational epitope and/or MEFA reacts specifically with anti-HCV antibodies present in a biological sample from an HCV-infected individual. The immunoassay of the invention is useful for detecting hepatitis C virus infection in a biological sample. The method of the invention provides a sensitive, accurate diagnostic and prognostic tool to provide adequate patient care and to



XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;  
 KW NS3/4a conformational epitope; multiple epitope fusion antigen; MEFA;  
 KW anti-HCV antibody; NS3/4a conformational antigen; HCV infection;  
 KW E2 hypervariable region.  
 XX  
 OS Hepatitis C virus type 1.  
 XX  
 PN US2002146685-A1.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 14-JUN-2001; 2001US-00881654.  
 XX  
 PR 15-JUN-2000; 2000US-0212082P.  
 PR 02-APR-2001; 2001US-0280811P.  
 PR 02-APR-2001; 2001US-0280867P.  
 XX  
 PA (CHIE/) CHIEN D Y.  
 PA (ARCA/) ARCANDEL P.  
 PA (TAND/) TANDESKE L.  
 PA (GEOR/) GEORGE-NASCIMENTO C.  
 PA (COIT/) COIT D.  
 PA (MEDI/) MEDINA-SELBY A.  
 XX  
 PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
 PI Medina-Selby A;  
 XX  
 DR WPI; 2003-147573/14.  
 XX  
 PT Immunoassay solid support for detecting Hepatitis C Virus infection in  
 PT biological samples, comprises Hepatitis C Virus conformational epitope  
 PT and multiple epitope fusion antigen.  
 XX  
 PS Example 3; Page 17; 45pp; English.  
 XX  
 CC The present invention relates to immunoassays comprising Hepatitis C  
 CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
 CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
 CC multiple epitope fusion antigen react with anti-HCV antibodies present in  
 CC a biological sample from an HCV-infected individual. The immunoassays and  
 CC methods of the invention are useful for detecting HCV infection in a  
 CC biological sample. The inventive immunoassay solid support provides a  
 CC sensitive and reliable method for detecting early HCV seroconversion. The  
 CC assays can detect HCV infection caused by any six known genotypes of HCV.  
 CC The use of the multiple epitope fusion proteins decreases masking  
 CC problems, improves sensitivity in detecting antibodies by allowing a  
 CC greater number of epitopes on a unit area of substrate, and improves  
 CC selectivity. The present sequence representing HCV type 1 (HCV-1) NS4a  
 CC peptide is used in a protease enzyme activity assay in the examples of  
 CC the present invention  
 XX  
 SQ Sequence 23 AA;  
 XX  
 "Query Match 90.0%; Score 99; DB 6; Length 23;  
 Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 Db 1 KKGSVVIVGRIVLSGKPAIIPKK 23  
 XX  
 RESULT 13  
 ABU61521  
 ID ABU61521 standard; peptide; 23 AA.  
 XX  
 AC ABU61521;  
 XX  
 DT 08-AUG-2003 (first entry)  
 XX  
 DE Hepatitis C virus NS3 protease enzyme assay associated peptide #2.  
 XX

KW Hepatitis C virus; NS3 protease; hepatotropic; virucide;  
 KW antiinflammatory; serine protease inhibitor; hepatitis C virus;  
 KW NS3 Protease inhibitor; viral infection; elastase; chymotrypsin.  
 XX  
 OS Synthetic.  
 XX  
 PN US200217725-A1.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 28-OCT-2001; 2001US-00039317.  
 XX  
 PR 23-OCT-2000; 2000US-0242557P.  
 XX  
 PA (PRIE/) PRIESTLEY E S.  
 XX  
 PI Priestley ES;  
 XX  
 DR WPI; 2003-465950/44.  
 XX  
 PT New class of peptides are inhibitors of hepatitis C virus NS3 protein,  
 PT useful for treating e.g. hepatitis.  
 XX  
 PS Example 39; Page 39; 54pp; English.  
 XX  
 CC The invention describes novel peptide useful for treating viral  
 CC infections e.g. Hepatitis C virus (HCV) infections. The compounds have  
 CC HCV NS3 protease inhibitory selectivity over inhibition of elastase and  
 CC chymotrypsin. This is the amino acid sequence of a modified peptide  
 CC associated with an assay of NS3 protease activity  
 XX  
 SQ Sequence 23 AA;  
 XX  
 Query Match 90.0%; Score 99; DB 7; Length 23;  
 Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
 Db 1 KKGSVVIVGRIVLSGKPAIIPKK 23  
 XX  
 RESULT 14  
 ADC06772  
 ID ADC06772 standard; peptide; 23 AA.  
 XX  
 AC ADC06772;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE HCV NS4A peptide used to monitor protease enzyme activity.  
 XX  
 KW immunoassay solid support; HCV; NS3/4a; non-structural;  
 KW non-A, non-B hepatitis; NANB; NS4A; NS3 cofactor.  
 OS  
 OS Hepatitis C virus.  
 XX  
 PN US2002192639-A1.  
 XX  
 PD 19-DEC-2002.  
 XX  
 PF 14-JUN-2001; 2001US-00881239.  
 XX  
 PR 15-JUN-2000; 2000US-0212082P.  
 PR 02-APR-2001; 2001US-0280811P.  
 PR 02-APR-2001; 2001US-0280867P.  
 XX  
 PA (CHIE/) CHIEN D Y.  
 PA (ARCA/) ARCANDEL P.  
 PA (TAND/) TANDESKE L.  
 PA (GEOR/) GEORGE-NASCIMENTO C.  
 PA (COIT/) COIT D.  
 PA (MEDI/) MEDINA-SELBY A.

```

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Seiby A;
XX WPI; 2003-644609/61.
DR
XX
XX Immunossay solid support for detecting hepatitis C virus infection in
PT biological samples, comprises a hepatitis C virus anti-core antibody and
PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core
PT antibody.
XX
XX Example 3; Page 18; 40pp; English.
XX
CC The invention relates to a novel immunoassay solid support comprising at
CC least one hepatitis C virus (HCV) anti-core antibody and at least one
CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.
CC The system of the invention may be useful for detecting HCV infection in
CC a biological sample and for treating or detecting non-A, non-B hepatitis
CC (NANB hepatitis). The current sequence is that of the HCV NS4A (NS3
CC cofactor) peptide of the invention which was used to monitor protease
CC enzyme activity.
XX
SQ Sequence 23 AA;
Query Match 90.0%; Score 99; DB 7; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKGSVIVIGRIILSGRPAIVPRR 23
DB 1 KKGSVIVIGRIILSGRPAIVPRR 23
RESULT 15
AAR82856
ID AAR82856 standard; peptide; 34 AA.
XX
AC AAR82856;
XX
XX 25-MAR-2003 (revised)
DT 09-APR-1996 (first entry)
XX
XX NS3 serine protease cofactor.
XX
XX NS3; serine protease; hepatitis C virus; HCV; NS4A; therapy.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 34
FT /label= Abu
FT
XX
PN WO9522985-A1.
XX
XX 31-AUG-1995.
PD
XX
XX 14-FEB-1995; 95WO-IT000018.
PF
XX
XX 23-FEB-1994; 94IT-RM000092.
PR
XX
XX (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.
PA
XX
XX De Francesco R, Failla C, Tomei L;
PI
XX WPI; 1995-311381/40.
DR
XX
XX In vitro reproduction of hepatitis C virus NS3 protease activity - by
PT including the NS4A cofactor in the mixt., useful for screening cpds. that
PT inhibit NS3.
XX
XX Example 3; Page 19; 26pp; English.
PS
XX This sequence represents a cofactor of the Hepatitis C virus (HCV) NS3

```

```

CC serine protease (see AAR82854). This sequence is based on the C-terminal
CC region of the NS4A protein (see AAR82855). Optimal serine protease
CC activity is obtained when NS4A and NS3 are present in a 1:1 ratio. The
CC cleavage site between these two proteins on the HCV genome can be mutated
CC so that the components remain covalently bonded. These sequences are
CC included in a composition that can be used in an assay system. This assay
CC system can be used to select compounds that inhibit NS3 activity, e.g.
CC potential therapeutic agents. NS4A can be used for screening enzyme
CC inhibitors. (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 34 AA;
Query Match 84.5%; Score 93; DB 2; Length 34;
Best Local Similarity 95.2%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GSVIVIGRIILSGRPAIVPRR 23
DB 1 GSVIVIGRIILSGRPAIVPRR 21
Search completed: July 6, 2004, 09:30:51
Job time : 40 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:28:15 ; Search time 14 seconds  
(without alignments)  
158.029 Million cell updates/sec

Title: US-10-070-220-13  
Perfect score: 110  
Sequence: 1 KKGWIVGRIILSGRPAIVPDR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: Pirl:\*  
2: Pirl2:\*  
3: Pirl3:\*  
4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 93    | 84.5        | 3010   | 1 GNVVTC | genome polyprotein |
| 2          | 92    | 83.6        | 3010   | 1 A45573 | genome polyprotein |
| 3          | 92    | 83.6        | 3010   | 1 S18030 | genome polyprotein |
| 4          | 91    | 82.7        | 3010   | 1 GNVVTC | genome polyprotein |
| 5          | 89    | 80.9        | 3010   | 1 GNVVTC | genome polyprotein |
| 6          | 86    | 78.2        | 3011   | 1 S40770 | genome polyprotein |
| 7          | 85    | 77.3        | 876    | 2 P82219 | polypeptide - hepa |
| 8          | 85    | 77.3        | 3014   | 1 JCS620 | genome polyprotein |
| 9          | 83    | 75.5        | 492    | 2 P80326 | polyprotein - hepa |
| 10         | 83    | 75.5        | 716    | 2 JQ1366 | polyprotein - hepa |
| 11         | 83    | 75.5        | 3011   | 1 GNVVTC | genome polyprotein |
| 12         | 82    | 74.5        | 3011   | 1 GNVVTC | genome polyprotein |
| 13         | 66    | 60.0        | 209    | 2 PCL306 | genome polyprotein |
| 14         | 64    | 58.2        | 142    | 2 PCL307 | genome polyprotein |
| 15         | 60    | 45.5        | 624    | 2 JCS471 | regulatory protein |
| 16         | 50    | 45.5        | 625    | 2 S18420 | regulatory protein |
| 17         | 49    | 44.5        | 244    | 2 B30887 | probable N-hydroxy |
| 18         | 49    | 44.5        | 244    | 2 H85730 | probable N-hydroxy |
| 19         | 49    | 44.5        | 281    | 2 B64899 | N-hydroxyarylamine |
| 20         | 49    | 44.5        | 908    | 2 B69435 | signal-transducing |
| 21         | 49    | 44.5        | 3228   | 2 T11381 | hypothetical prote |
| 22         | 47    | 42.7        | 117    | 2 T4264  | hypothetical prote |
| 23         | 47    | 42.7        | 144    | 2 T29958 | hypothetical prote |
| 24         | 47    | 42.7        | 211    | 2 T10072 | periplasmic protei |
| 25         | 47    | 42.7        | 570    | 2 A75201 | hypothetical prote |
| 26         | 46.5  | 42.3        | 432    | 2 B95202 | probable cytosine  |
| 27         | 46    | 41.8        | 306    | 2 T51705 | methylenetetrahydr |
| 28         | 46    | 41.8        | 464    | 2 AGC280 | NAD(P) transhydrog |
| 29         | 45.5  | 41.4        | 93     | 2 C55855 | holin homolog - ph |

|    |      |      |     |          |                     |
|----|------|------|-----|----------|---------------------|
| 30 | 45   | 40.9 | 94  | 2 T03376 | dihydroliipoamide S |
| 31 | 45   | 40.9 | 117 | 2 T24263 | hypothetical prote  |
| 32 | 45   | 40.9 | 143 | 2 T24261 | hypothetical prote  |
| 33 | 45   | 40.9 | 256 | 2 A95100 | ABC transporter, A  |
| 34 | 45   | 40.9 | 263 | 2 C97968 | hypothetical prote  |
| 35 | 45   | 40.9 | 281 | 2 T22280 | hypothetical prote  |
| 36 | 45   | 40.9 | 480 | 2 S77206 | NAD(P) transhydrog  |
| 37 | 45   | 40.9 | 570 | 2 E71234 | hypothetical prote  |
| 38 | 44   | 40.0 | 104 | 2 H87548 | hypothetical prote  |
| 39 | 44   | 40.0 | 141 | 2 T27071 | hypothetical prote  |
| 40 | 44   | 40.0 | 179 | 2 D90167 | conserved hypotet   |
| 41 | 44   | 40.0 | 230 | 2 B56122 | emopamil-binding p  |
| 42 | 44   | 40.0 | 303 | 2 H75528 | conserved hypotet   |
| 43 | 44   | 40.0 | 315 | 2 A71288 | probable methylene  |
| 44 | 43.5 | 39.5 | 88  | 2 D95022 | conserved hypotet   |
| 45 | 43.5 | 39.5 | 88  | 2 G97893 | hypothetical prote  |

## ALIGNMENTS

## RESULT 1

GNVVTG

genome polyprotein - hepatitis C virus

N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001

C;Accession: A38465

R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;  
J. Virol. 65, 1105-1113, 1991

A;Title: Structure and organization of the hepatitis C virus genome isolated from human

A;Reference number: A38465; MUID:91140698; PMID:1847440

A;Accession: A38465

A;Molecule type: genomic RNA

A;Residues: 1-3010 &lt;TAX&gt;

A;Cross-references: EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G329771

C;Superfamily: hepatitis C virus genome polyprotein; glycoprotein; hydrolase; nonstructura  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; glycoprotein; hydrolase; nonstructura

F;2-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F;116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F;192-389/Product: major envelope protein E #status predicted &lt;MEE&gt;

F;390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F;730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F;1007-1635/Product: hepacivirin #status predicted &lt;NS3&gt;

F;1230-1237/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEHX motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;

F;1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;

F;2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F;1396,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 84.5%; Score 93; DB 1; Length 3010;

Best Local Similarity 95.2%; Pred. No. 1.6e-05;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVIVIGRIILSGRPAIVPDR 23

Db 1678 GSVIVIGRIILSGRPAIVPDR 1698

## RESULT 2

A45573

genome polyprotein - hepatitis C virus (strain JT)

N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

C;Accession: A45573

R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,

Virus Res. 23, 39-53, 1992

A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 83.6%; Score 92; DB 1; Length 3010;  
Best Local Similarity 90.5%; Pred. No. 2.3e-05;  
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPR 23  
|||||:|||||:  
Db 1678 GSVVIVGRILSGRPAIVPR 1698

RESULT 4  
GNWVCJ  
genome polyprotein - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
C:Accession: A39253; PS0086  
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, N.; Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A:Reference number: A39253; MUID:91088550; PMID:2175903  
A:Accession: A39253  
A:Molecule type: Genomic RNA  
A:Residues: 1-3010 <KAT>  
A:Cross-references: GB:D90208; NID:Q221610; PIDN:BAAL4233.1; PID:Q221611  
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari  
A:Reference number: PS0085  
A:Accession: PS0086  
A:Molecule type: Genomic RNA  
A:Residues: 2650-2707 <KAT>  
A:Experimental source: Japanese isolate  
C:Comment: The cleavage sites of this polyprotein have not been determined.  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; Glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:116-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>  
F:1230-1337/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,417,423,448,532,556,576,623,645,1213,1255,2041,2077,2240,2,2

Query Match 82.7%; Score 91; DB 1; Length 3010;  
Best Local Similarity 85.7%; Pred. No. 3.3e-05;  
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPR 23  
|||||:|||||:  
Db 1678 GSVVIVGRILSGRPAIVPR 1698

RESULT 5  
GNWVTW  
genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40244



R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992  
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A;Reference number: A40244; MUID:92230206; PMID:1314449  
A;Accession: A40244  
A;Molecule type: Genomic RNA  
A;Residues: 1-3010 <CHE>  
A;Cross-references: GB:M84754  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F;11-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-725/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polypeptide <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 207

Query Match 80.9%; Score 89; DB 1; Length 3010;  
Best Local Similarity 85.7%; Pred. No. 6.7e-05;  
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVIVGRIILSGRPAIVPRR 23  
Db 1678 GSVIVGRIILSGRPAIVPRR 1698

RESULT 6  
S40770  
genome polypeptide - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C;Accession: S40770; PC1285  
R;Okamoto, H.  
Submitted to the EMBL Data Library, March 1992  
A;Reference number: S40770  
A;Accession: S40770  
A;Molecule type: Genomic RNA  
A;Residues: 1-3011 <ORA>  
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587  
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,  
Jpn. J. Exp. Med. 60, 167-177, 1990  
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A;Reference number: PC1284; MUID:91013116; PMID:2170712  
A;Accession: PC1285  
A;Molecule type: Genomic RNA  
A;Residues: 1-513 <OK>  
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512  
A;Experimental source: isolate HC-J1  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F;116-191/Product: capsid protein C #status predicted <CPC>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-725/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polypeptide <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 78.2%; Score 86; DB 1; Length 3011;

Best Local Similarity 81.0%; Pred. No. 0.00019;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GSVIVGRIILSGRPAIVPRR 23  
Db 1678 GSVIVGRIILSGRPAIVPRR 1698

RESULT 7  
PC2219  
polypeptide - hepatitis C virus (type 5a) (fragments)  
N;Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote  
C;Species: hepatitis C virus  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Nov-2000  
C;Accession: PC2219  
R;Stuyver, L.; Arnheim, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A;Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the  
A;Reference number: PC2219; MUID:94338342; PMID:7520237  
A;Accession: PC2219  
A;Molecule type: mRNA  
A;Residues: 1-876 <STU>  
A;Cross-references: GB:L29577; GB:L29578; GB:L29579  
A;Experimental source: serum  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: Glycoprotein  
F;1-191/Product: core #status predicted <COE>  
F;68-78/Region: variable  
F;192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F;248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F;248-338/Region: E2  
F;339-411/Region: NS1 (amino end)  
F;412-783/Product: NS3 #status predicted <NSR>  
F;784-837/Product: NS4A #status predicted <NSA>  
F;838-876/Product: NS4B #status predicted <NSB>  
F;281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 77.3%; Score 85; DB 2; Length 876;  
Best Local Similarity 81.0%; Pred. No. 7.7e-05;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GSVIVGRIILSGRPAIVPRR 23  
Db 804 GSVIVGRIILSGRPAIVPRR 824

RESULT 8  
JC5620  
genome polypeptide - hepatitis C virus (isolate EUH1480)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C;Accession: JC5620  
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
A;Reference number: JC5620; MUID:97366593; PMID:9223423  
A;Accession: JC5620  
A;Molecule type: mRNA  
A;Residues: 1-3014 <CHA>  
A;Cross-references: GB:Y13184  
A;Experimental source: genotype 5a, which predominates in South Africa  
A;Note: the translation of the nucleotide sequence is not complete in this paper  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F;116-191/Product: capsid protein C #status predicted <CPC>  
F;192-389/Product: major envelope protein E #status predicted <EPM>  
F;384-408/Region: hypervariable #status predicted <MEE>  
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1008-1616/Product: hepatitis C virus genome polypeptide <NS3>

F:1231-1338/Region: nucleotide-binding motif A (P-loop)  
 F:1313-1318/Region: nucleotide-binding motif B  
 F:1317-1320/Region: DEXH motif  
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 77.3%; Score 85; DB 1; Length 3014;  
 Best Local Similarity 85.7%; Pred. No. 0.00027;  
 Matches 18; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
 DB 1679 GSAIVGRILSGRPAIVPDR 1699

RESULT 9  
 PS0326  
 C:Species: hepatitis C virus (isolate Fla) (fragments)  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
 C:Accession: PS0326  
 R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.  
 Gene 105, 167-172, 1991  
 A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype  
 A:Reference number: PS0326; MUID:92039028; PMID:1718820  
 A:Accession: PS0326  
 A:Molecule type: Genomic RNA  
 A:Residues: 1-492 <LIJ>  
 A:Cross-references: GB:M60220  
 A:Note: this sequence corresponds to nonstructural protein NS3 region  
 A:Note: translation of the nucleotide sequence is not complete  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: polyprotein

Query Match 75.5%; Score 83; DB 2; Length 492;  
 Best Local Similarity 76.2%; Pred. No. 8.7e-05;  
 Matches 16; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
 DB 229 GCWVIVGRILSGKPAIIPDR 249

RESULT 10  
 JQ1366  
 C:Species: hepatitis C virus (French isolate) (fragments)  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
 C:Accession: JQ1366  
 R:Kremsdorff, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.  
 J. Gen. Virol. 72, 2557-2561, 1991  
 A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication  
 A:Reference number: JQ1366; MUID:92013977; PMID:1655961  
 A:Accession: JQ1366  
 A:Molecule type: Genomic RNA  
 A:Residues: 1-716 <kRE>  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: glycoprotein; polyprotein  
 F:84, 90, 97, 115, 143, 199, 223, 243, 290, 312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 75.5%; Score 83; DB 2; Length 716;  
 Best Local Similarity 76.2%; Pred. No. 0.00013;  
 Matches 16; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
 DB 627 GCWVIVGRILSGKPAIIPDR 647

RESULT 11  
 GNVVCH

genome polyprotein - hepatitis C virus (strain H)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A36814; A41546  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to GenBank, July 1992  
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: comparison of reference number: A36814  
 A:Accession: A36814  
 A:Molecule type: Genomic RNA  
 A:Residues: 1-3011 <INC>  
 A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison of reference number: A41546; MUID:92052256; PMID:1658800  
 A:Contents: annotation  
 A:Note: neither amino acid nor nucleotide sequence is given  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 F:115-191/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEP>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1635/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 23

Query Match 75.5%; Score 83; DB 1; Length 3011;  
 Best Local Similarity 76.2%; Pred. No. 0.00055;  
 Matches 16; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
 DB 1678 GCWVIVGRILSGKPAIIPDR 1698

RESULT 12  
 GNVVCH  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C:Accession: A39166; PQ0403  
 R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A:Title: Genetic organization and diversity of the hepatitis C virus.  
 A:Reference number: A39166; MUID:91172826; PMID:1848704  
 A:Accession: A39166  
 A:Molecule type: mRNA  
 A:Residues: 1-3011 <CHO>  
 A:Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
 A:Accession: PQ0403  
 A:Molecule type: Genomic RNA  
 A:Residues: 1577-1633 <CHA>  
 A:Cross-references: DDBJ:D10128  
 A:Experimental source: isolates E-b16  
 A:Accession: PQ0404  
 A:Status: preliminary

A:Molecule type: Genomic RNA  
A:Residues: 1577-1633 <CH2>  
A:Experimental source: Isolates E-b17  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 74.5%; Score 82; DB 1; Length 3011;  
Best Local Similarity 71.4%; Pred. No. 0.00077;  
Matches 15; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GSWIVGRIILSGRPAIVP 23  
Db 1678 GCVIVGRVLSGKPAIVP 1698  
|||||:|||||:|||||

RESULT 13  
PC1306  
genome polyprotein NS4a epitope containing region (isolate BR36-20) - hepatitis C virus  
C:Species: hepatitis C virus  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 17-Nov-2000  
C:Accession: PC1306  
R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; Deleys, R.; Maertens, G.  
Biochem. Biophys. Res. Commun. 192, 635-641, 1993  
A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3.  
A:Reference number: PC1300; MUID:93249436; PMID:7683463  
A:Accession: PC1306  
A:Molecule type: mRNA  
A:Residues: 1-209 <STU>  
A:Cross-references: DBJ:DL4600; NID:G303584; PIDN:BAA03449.1; PID:G303585  
A:Experimental source: blood  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: nonstructural protein; polyprotein

Query Match 60.0%; Score 66; DB 2; Length 209;  
Best Local Similarity 73.7%; Pred. No. 0.014;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GSWIVGRIILSGRPAIVP 21  
Db 123 GCVIVGHIELGKPAIVP 141  
|||||:|||||:|||||

RESULT 14  
PC1307  
genome polyprotein NS4a epitope containing region (isolate HD10-1) - hepatitis C virus  
C:Species: hepatitis C virus  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 17-Nov-2000  
C:Accession: PC1307  
R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; Deleys, R.; Maertens, G.  
Biochem. Biophys. Res. Commun. 192, 635-641, 1993  
A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3.  
A:Reference number: PC1300; MUID:93249436; PMID:7683463  
A:Accession: PC1307  
A:Molecule type: mRNA  
A:Residues: 1-142 <STU>  
A:Cross-references: DBJ:DL4602  
A:Experimental source: blood  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: polyprotein

Query Match 58.2%; Score 64; DB 2; Length 142;  
Best Local Similarity 68.4%; Pred. No. 0.019;  
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GSWIVGRIILSGRPAIVP 21  
Db 56 GCVIVGHIELGKPAIVP 74  
|||||:|||||:|||||

RESULT 15  
JC5471  
regulatory protein nifa - Azospirillum lipoferum  
C:Species: Azospirillum lipoferum  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 28-Apr-2003  
C:Accession: JC5471  
R:Shigenatsu, T.; Inoue, A.; Hidaka, M.; Masaki, H.; Uozumi, T.  
Biosci. Biotechnol. Biochem. 61, 768-771, 1997  
A:Title: Oxygen sensitivity of NifA protein of Azospirillum lipoferum FS as suggested by  
A:Reference number: JC5471; MUID:97321848; PMID:9178550  
A:Accession: JC5471  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-624 <SHI>  
A:Cross-references: DDBJ:DJ3799; NID:G1384045; PIDN:BAA02956.1; PID:G216222  
A:Experimental source: strain FS  
A:Note: submitted to the DDBJ/EMBL/GenBank databases December, 1992  
C:Comment: This protein activates the nifH promoter under microaerobic conditions.  
C:Genetics:  
A:Gene: nifa  
C:Superfamily: Response regulator (sigma54-dependent transcriptional activator), FlhA ty  
C:Keywords: DNA binding; iron-sulfur protein; metalloprotein; P-loop; transcription reg  
F:203-424/Domain: RNA polymerase sigma factor interaction domain homology <SFI>  
F:231-238/Region: nucleotide-binding motif A (P-loop) #status atypical  
F:298-302/Region: nucleotide-binding motif B  
F:582-624/Domain: DNA-binding #status predicted <DNA>  
F:596-615/Region: helix-turn-helix  
F:413,425,445,450/Binding site: iron-sulfur clusters (Cys) (covalent) #status predicted

Query Match 45.5%; Score 50; DB 2; Length 624;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 7 IVGRRIILSGRPAIVP 21  
Db 85 ITGRILKTGMFPAIVP 99  
|||||:|||||:|||||

Search completed: July 6, 2004, 09:28:59  
Job time : 15 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:28:14 ; Search time 9 Seconds  
(without alignments)  
133.068 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110

Sequence: 1 KKGSVIVIGRIILSRPAIVPR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 93    | 84.5        | 3010   | 1 POLG_HCVBK | P26663 h genome po |
| 2          | 92    | 83.6        | 3010   | 1 POLG_HCVUT | Q00269 h genome po |
| 3          | 91    | 82.7        | 3010   | 1 POLG_HCVUA | P26662 h genome po |
| 4          | 89    | 80.9        | 3010   | 1 POLG_HCVTV | P29846 h genome po |
| 5          | 83    | 75.5        | 3011   | 1 POLG_HCVH  | P27958 h genome po |
| 6          | 82    | 74.5        | 3011   | 1 POLG_HCVI  | P26664 h genome po |
| 7          | 50    | 45.5        | 624    | 1 NIFA_AZOLI | P54929 azospirilla |
| 8          | 50    | 45.5        | 625    | 1 NIFA_AZOR  | P30667 azospirilla |
| 9          | 49    | 44.5        | 281    | 1 NHOA_ECOLI | P77567 escherichia |
| 10         | 48    | 43.6        | 334    | 1 PLXK_CLOTE | Q895n0 clostridium |
| 11         | 47    | 42.7        | 211    | 1 MAUD_METME | Q50232 methylophil |
| 12         | 46    | 41.8        | 248    | 1 ISPD_COREF | Q8fm33 corynebacte |
| 13         | 46    | 41.8        | 332    | 1 PLXK_THETF | Q8r9v6 thermococae |
| 14         | 45.5  | 41.4        | 93     | 1 HOLI_BPP2  | P51773 bacterioph  |
| 15         | 45    | 40.9        | 523    | 1 C756_CAMME | O04773 campanula m |
| 16         | 44.5  | 40.5        | 338    | 1 G3P_PPHAR  | O13507 phaffia rho |
| 17         | 44    | 40.0        | 179    | 1 Y256_SUSJO | Q990n5 sulfobolus  |
| 18         | 44    | 40.0        | 230    | 1 EBP_HUMAN  | Q15125 homo sapien |
| 19         | 43.5  | 39.5        | 885    | 1 YDGH_BACSU | P96706 bacillus su |
| 20         | 43    | 39.1        | 339    | 1 PLXK_CLOPE | Q8xjng clostridium |
| 21         | 42.5  | 38.6        | 411    | 1 Y867_METTH | O69553 methanobact |
| 22         | 42    | 38.2        | 138    | 1 Y661_SULTO | Q976x8 sulfobolus  |
| 23         | 42    | 38.2        | 224    | 1 YNFK_ECOLI | P75974 escherichia |
| 24         | 42    | 38.2        | 305    | 1 LIGD_PSEPA | Q01198 pseudomonas |
| 25         | 42    | 38.2        | 333    | 1 PLXK_CLOTS | O65984 clostridium |
| 26         | 42    | 38.2        | 462    | 1 PNTE_ECOLI | P07002 escherichia |
| 27         | 42    | 38.2        | 594    | 1 NIFA_RHISN | Q53206 rhizobium s |
| 28         | 42    | 38.2        | 736    | 1 TREE_MAGGR | O42622 magnaporthe |
| 29         | 42    | 38.2        | 1295   | 1 PUR4_SALTI | Q8z416 salmonella  |
| 30         | 42    | 38.2        | 1295   | 1 PUR4_SALTI | P74881 salmonella  |
| 31         | 41.5  | 37.7        | 457    | 1 TRME_SHEON | Q8cx52 shewanella  |
| 32         | 41    | 37.3        | 141    | 1 YQ07_CABEL | P34645 caenorhabdi |
| 33         | 41    | 37.3        | 256    | 1 Y487_THEAC | Q9hk66 thermoplasm |

#### ALIGNMENTS

##### RESULT 1

##### POLG\_HCVBK

ID POLG\_HCVBK STANDARD; PRT; 3010 AA.

AC P26663;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate BX) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=11105;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91140698; PubMed=1847440;

RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,

Onishi E., Andon T., Yoshida I., Okayama H.;

"Structure and organization of the hepatitis C virus genome isolated

from human carriers.";

J. Virol. 65:1105-1113(1991).

[2]

SEQUENCE OF 1487-1500.

MEDLINE=96235224; PubMed=8647104;

Borowski P., Helland M., Oehlmann K., Becker B., Kornetevy L.;

"Non-structural protein 3 of hepatitis C virus inhibits

phosphorylation mediated by cAMP-dependent protein kinase.";

Eur. J. Biochem. 237:611-618(1996).

[3]

X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.

MEDLINE=97015088; PubMed=861916;

Love R.A., Farge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,

Moomaw E.W., Adachi T., Hostomsky Z.;

"The crystal structure of hepatitis C virus NS3 proteinase reveals a

trypsin-like fold and a structural zinc binding site.";

Cell 87:331-342(1996).

[4]

X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.

MEDLINE=98227846; PubMed=9568891;

Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,

Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;

"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C

virus: a 2.2-A resolution structure in a hexagonal crystal form.";

Protein Sci. 7:837-847(1998).

CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are

hydrophobic, suggesting a possible membrane-related function. NS3

and NS5 may play a role in the viral RNA replication.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

precursor polyprotein, commonly with Asp or Glu in the P6

position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +



Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).

OS Hepatitis C virus (isolate HC-JT) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI\_TaxID=31642;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92295714; PubMed=1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;

RA "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals.";

RT Virus Res. 23:39-53(1992).

RL

CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D11168; BA01943.1; -.

DR F01; A45573; A45573.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR InterPro; IPR009003; Cys Ser\_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV NS5b.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA pol DS PS.

DR InterPro; IPR007094; RNA pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01003; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RGRP; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDc; 1.

KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolyase; Serine protease. INT\_MET 1 1

FT CHAIN 1 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 729

FT CHAIN 730 1006

FT CHAIN 1007 1615

FT CHAIN 1616 1862

FT CHAIN 1863 2013

FT CHAIN 2014 3010

FT TRANSMEM 347 369

FT ACT\_SITE 1083 1083

FT ACT\_SITE 1107 1107

FT ACT\_SITE 1165 1165

FT NP\_BIND 1230 1237

FT SITE 1316 1319

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

FT CARBOHYD 250 250

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 532 532

FT CARBOHYD 540 540

FT CARBOHYD 556 556

FT CARBOHYD 576 576

FT CARBOHYD 623 623

FT CARBOHYD 645 645

FT CARBOHYD 2041 2041

FT CARBOHYD 2077 2077

FT CARBOHYD 2240 2240

FT CARBOHYD 2529 2529

FT CARBOHYD 2788 2788

SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 83.6%; Score 92; DB 1; Length 3010;

Best Local Similarity 90.5%; Pred. No. 1.2e-05;

Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPR 23

DB 1678 GSVVIVGRILSGRPAIVPR 1698

RESULT 3

POLG\_HCVJA STANDARD; PRT; 3010 AA.

ID POLG\_HCVJA

AC P26662;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Japanese) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI\_TaxID=111116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91088550; PubMed=2175903;

RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,

Sugimura T., Shimotohno K.;  
"Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis";  
Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
(2)  
RN  
DISCUSSION OF SEQUENCE.  
MEDLINE=91192160; PubMed=1849488;  
Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,  
Okoshi S., Shimotohno K.;  
"Molecular structure of the Japanese hepatitis C viral genome.";  
FEBS Lett. 280:325-328(1991).  
CC  
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
hydrophobic, suggesting a possible membrane-related function. NS3  
and NS5 may play a role in the viral RNA replication.  
CC  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
precursor polyprotein, commonly with Asp or Glu in the P6  
position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
{RNA} (N).  
CC  
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
lipoprotein envelope. The envelope consists of two proteins:  
protein M and glycoprotein E. The nucleocapsid is a complex of  
protein C and mRNA.  
CC  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----

EMBL; D90208; BAA14233.1; -.  
DR  
PIR; A39253; GNWVCU.  
DR  
HSSP; P26663; LUXP.  
DR  
MEROPS; S29.001; -.  
DR  
MEROPS; U39.001; -.  
DR  
InterPro; IPR009003; Cys\_ser\_trypsin.  
DR  
InterPro; IPR001410; DEAD.  
DR  
InterPro; IPR002522; HCV\_capsid.  
DR  
InterPro; IPR002521; HCV\_core.  
DR  
InterPro; IPR002519; HCV\_env.  
DR  
InterPro; IPR002531; HCV\_NS1.  
DR  
InterPro; IPR002518; HCV\_NS2.  
DR  
InterPro; IPR000745; HCV\_NS4a.  
DR  
InterPro; IPR001490; HCV\_NS4b.  
DR  
InterPro; IPR002868; HCV\_NS5a.  
DR  
InterPro; IPR002166; HCV\_RdRp.  
DR  
InterPro; IPR001650; Helicase\_C.  
DR  
InterPro; IPR004103; Peptidase\_C29.  
DR  
InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR  
InterPro; IPR007094; RNA\_pol\_PSVir.  
DR  
Pfam; PF01543; HCV\_capsid; 1.  
DR  
Pfam; PF01542; HCV\_core; 1.  
DR  
Pfam; PF01539; HCV\_env; 1.  
DR  
Pfam; PF01560; HCV\_NS1; 1.  
DR  
Pfam; PF01538; HCV\_NS2; 1.  
DR  
Pfam; PF02907; HCV\_NS3; 1.  
DR  
Pfam; PF01006; HCV\_NS4a; 1.  
DR  
Pfam; PF01001; HCV\_NS4b; 1.  
DR  
Pfam; PF01506; HCV\_NS5a; 1.  
DR  
Pfam; PF00271; Helicase\_C; 1.  
DR  
Pfam; PF00998; Viral\_RdRp; 1.  
DR  
ProDom; PD186062; HCV\_NS1; 1.  
DR  
SMART; SM00487; DEXDC1; 1.  
DR

Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
INIT\_MET 1 1  
CHAIN 116 191  
FT CHAIN  
FT CAPSID PROTEIN C (POTENTIAL) .  
FT MATRIX PROTEIN (POTENTIAL) .  
FT CELLULAR AMINOPEPTIDASE.  
FT REMOVED FROM CAPSID PROTEIN C BY THE

|     |          |          |                                    |  |
|-----|----------|----------|------------------------------------|--|
| FT  | CHAIN    | 192      | 383                                | MAJOR ENVELOPE PROTEIN E (POTENTIAL).    |
| FT  | CHAIN    | 384      | 729                                | NON-STRUCTURAL PROTEIN NS1 (POTENTIAL).  |
| FT  | CHAIN    | 730      | 1006                               | NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).  |
| FT  | CHAIN    | 1007     | 1615                               | PROTEASE/HELICASE NS3 (POTENTIAL).       |
| FT  | CHAIN    | 1616     | 1662                               | NON-STRUCTURAL PROTEIN NS4A (POTENTIAL). |
| FT  | CHAIN    | 1663     | 2013                               | NON-STRUCTURAL PROTEIN NS4B (POTENTIAL). |
| FT  | CHAIN    | 2014     | 3010                               | RNA-DIRECTED RNA POLYMERASE (POTENTIAL). |
| FT  | TRANSMEM | 347      | 369                                | POTENTIAL.                               |
| FT  | ACT_SITE | 1083     | 1083                               | CHARGE RELAY SYSTEM (BY SIMILARITY).     |
| FT  | ACT_SITE | 1107     | 1107                               | CHARGE RELAY SYSTEM (BY SIMILARITY).     |
| FT  | ACT_SITE | 1165     | 1165                               | CHARGE RELAY SYSTEM (BY SIMILARITY).     |
| FT  | NP_BIND  | 1230     | 1237                               | ATP (POTENTIAL).                         |
| FT  | SITE     | 1316     | 1319                               | DECH BOX.                                |
| FT  | CARBOHYD | 196      | 196                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 209      | 209                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 234      | 234                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 250      | 250                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 305      | 305                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 417      | 417                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 423      | 423                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 430      | 430                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 448      | 448                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 532      | 532                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 556      | 556                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 576      | 576                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 623      | 623                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 645      | 645                                | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 2041     | 2041                               | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 2077     | 2077                               | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 2240     | 2240                               | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| FT  | CARBOHYD | 2788     | 2788                               | N-LINKED (GLCNAC. .) (POTENTIAL).        |
| SEQ | SEQUENCE | 3010 AA; | 327017 MW; AA993794F46DB185 CRC64; |  |

Query Match 82.7%; Score 91; DB 1; Length 3010;

Best Local Similarity 85.7%; Pred. No. 1.8e-05;

Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps

QY 3 GSVIVGRILLSGRPAIVPRR 23

DB 1678 GSVIVGRILLSGRPAIVPDR 1698

|||||:::|

RESULT 4

POLG\_HCVTW

|     |   |           |               |
|-----|---|-----------|---------------|
| AC  | POLG_HCVTW  | STANDARD; | PRT; 3010 AA. |
| ID  | P29846;   |           |               |
| DD  | 01-APR-1993 (Rel. 25, Created)  |           |               |
| DT  | 01-APR-1993 (Rel. 25, Last sequence update)   |           |               |
| DT  | 10-OCT-2003 (Rel. 42, Last annotation update)   |           |               |
| DT  | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. |           |               |
| DE  | Hepatitis C virus (isolate Taiwan) (HCV).   |           |               |
| OS  | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  |           |               |
| OC  | NCHI_taxid=31645;   |           |               |
| OX  | [1]   |           |               |
| RP  | SEQUENCE FROM N.A.  |           |               |
| RR  | MEDLINE=92230206; PubMed=1314449;   |           |               |
| RX  | Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.; "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' terminus of viral genomic and antigenomic RNA.;"   |           |               |
| RT  | Virology 188:102-113(1992).   |           |               |
| RRL | -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  |           |               |
| CCC | -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.   |           |               |
| CCC |   |           |               |







FT HELIX 1555 1564  
 FT HELIX 1570 1578  
 FT TURN 1579 1580  
 FT TURN 1584 1597  
 FT TURN 1598 1598  
 FT HELIX 1606 1611  
 FT TURN 1614 1618  
 FT STRAND 1622 1623  
 FT STRAND 1627 1627  
 FT STRAND 1635 1636  
 FT HELIX 1640 1652  
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CDD94753 CRC64;

Query Match 75.58; Score 83; DB 1; Length 3011;  
 Best Local Similarity 76.23; Pred. No. 0.00029;

Matches 16; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSVVIVGRILISGRPAIVPRR 23

Db 1678 GCVVIVGRIVLSGKPAIIPR 1698

RESULT 6  
 ID POLG\_HCV1 STANDARD; PRT; 3011 AA.  
 AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.-98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate 1) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.

OC NCBI\_TaxID=11104;

OX [1]

RN SEQUENCE FROM N.A.

RX MEDLINE-91172826; PubMed-1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Kuo G., Houghton M.,

RT "Genetic organization and diversity of the hepatitis C virus.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are

CC hydrophobic, suggesting a possible membrane-related function. NS3

CC and NS5 may play a role in the viral RNA replication.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC (RNA) (N).

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a

CC lipoprotein envelope. The envelope consists of two proteins:

CC protein M and glycoprotein E. The nucleocapsid is a complex of

CC protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M62321; AA45676.1; -

CC FIR; A39166; GNWVC3.

CC PDB; 1a1v; 16-FEB-99.

DR PDB; 1HEI; 25-NOV-98.  
 DR MEROPS; S29.001; -.  
 DR MEROPS; U39.001; -.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRP.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; Peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_Psvit.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral RdRP; 1.  
 DR ProDom; PD186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 DR 3D-structure.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match. 74.5%; Score 82; DB 1; Length 3011;

Best Local Similarity 71.4%; Pred. No. 0.0004;

Matches 15; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPR 23

Db 1678 GCVVIVGRVLSGKPAIIDR 1698

# RESULT 7

NIFA AZOLI  
ID NIFA AZOLI STANDARD; PRT; 624 AA.  
AC PS4929;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Nif-specific regulatory protein.  
GN NIFA.  
OS Azospirillum lipoferum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Rhodospirillaceae; Azospirillum.  
OX NCBI\_TaxID=193;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shigematsu T., Hidaka M., Masaki H., Uozumi T.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NIFA, A TRANSCRIPTIONAL ACTIVATOR, IS REQUIRED FOR  
CC ACTIVATION OF MOST NIF OPERONS, WHICH ARE DIRECTLY INVOLVED IN  
CC NITROGEN FIXATION. NIFA INTERACTS WITH SIGMA-54.  
CC -!- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding  
CC domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

EMBL; D13799; BAA02956.1; --  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR008931; FIS-like.  
DR InterPro; IPR003018; GAF.  
DR InterPro; IPR002197; HTH\_Fis.  
DR InterPro; IPR002078; Sig54\_interact.  
DR Pfam; PF01590; GAF; 1.  
DR Pfam; PF02954; HTH\_8; 1.  
DR PRINTS; PF01590; Sig54\_activat; 1.  
DR PRINTS; PF01590; HTHFIS.  
DR SMART; SM00382; AAA; 1.  
DR SMART; SM00065; GAF; 1.  
DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
KW Nitrogen fixation; Transcription regulation; Activator;  
KW ATP-binding; DNA-binding; Metal-binding.  
FT DOMAIN 203 431 SIGMA-54 FACTOR INTERACTION (POTENTIAL).  
FT DOMAIN 432 581 INTER-DOMAIN LINKER.  
FT NP\_BIND 231 238 C-TERMINAL DNA-BINDING DOMAIN.  
FT NP\_BIND 294 303 ATP (POTENTIAL).  
FT METAL 445 445 BY SIMILARITY.  
FT METAL 450 450 BY SIMILARITY.  
FT DNA\_BIND 596 615 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 624 AA; 67119 MW; 223A36AD44B7016 CRC64;

Query Match 45.5%; Score 50; DB 1; Length 624;

Best Local Similarity 60.0%; Pred. No. 6.3;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 IVGRILSGRPAIVP 21

Db 85 ITGRILKTMGPAVVP 99

# RESULT 8

NIFA AZOBR  
ID NIFA AZOBR STANDARD; PRT; 625 AA.  
AC PS0867;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Nif-specific regulatory protein.  
GN NIFA.  
OS Azospirillum brasilense.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Rhodospirillaceae; Azospirillum.  
OX NCBI\_TaxID=192;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Liang Y.Y., Kaminski P.A., Elmerich C.;  
RL MEDLINE=92140038; PubMed=1779763;  
CC -!- IDENTIFICATION OF A NIFA-LIKE REGULATORY GENE OF AZOSPIRILLUM  
CC brasilense Sp7 expressed under conditions of nitrogen fixation and in  
CC the presence of air and ammonia.  
CC Mol. Microbiol. 5:2735-2744(1991).  
RN [2]  
RP CHARACTERIZATION.  
RA Liang Y.Y., de Zamaroczy M., Arsene F., Paquelin A., Elmerich C.;  
RL MEDLINE=93122522; PubMed=1362170;  
CC -!- Regulation of nitrogen fixation in Azospirillum brasilense Sp7;  
CC involvement of nifA, glnA and glnB gene products.  
CC FEMS Microbiol. Lett. 79:113-119(1992).  
CC -!- FUNCTION: NIFA, A TRANSCRIPTIONAL ACTIVATOR, IS REQUIRED FOR  
CC ACTIVATION OF MOST NIF OPERONS, WHICH ARE DIRECTLY INVOLVED IN  
CC NITROGEN FIXATION. NIFA INTERACTS WITH SIGMA-54.  
CC -!- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding  
CC domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

EMBL; X60714; CAA43126.1; --  
DR FIR; S18420; S18420.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR008931; FIS-like.  
DR InterPro; IPR003018; GAF.  
DR InterPro; IPR002197; HTH\_Fis.  
DR InterPro; IPR002078; Sig54\_interact.  
DR Pfam; PF01590; GAF; 1.  
DR Pfam; PF02954; HTH\_8; 1.  
DR PRINTS; PF01590; Sig54\_activat; 1.  
DR PRINTS; PF01590; HTHFIS.  
DR SMART; SM00382; AAA; 1.  
DR SMART; SM00065; GAF; 1.  
DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
KW Nitrogen fixation; Transcription regulation; Activator;  
KW ATP-binding; DNA-binding.  
FT DOMAIN 205 433 SIGMA-54 FACTOR INTERACTION (POTENTIAL).  
FT DOMAIN 434 582 INTER-DOMAIN LINKER.  
FT DOMAIN 583 625 C-TERMINAL DNA-BINDING DOMAIN.  
FT NP\_BIND 233 240 ATP (POTENTIAL).

```

DR Pfam; PF00797; Acetyltransferase2; 1.
DR PRINTS; PR01543; ANATRNSPASE.
DR ProDom; PD002783; Acetyltransferase2; 1.
DR Transferrase; Acyltransferase; Complete proteome.
FT ACT SITE 69 BY SIMILARITY.
SQ SEQUENCE 281 AA; 32274 MW; D6B777E05B629D2 CRC64;

Query Match 44.5%; Score 49; DB 1; Length 281;
Best Local Similarity 50.0%; Pred. No. 4.2;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 7 IVGRITLSGRPAIVPR 22
   :|::|||::|:
Db 90 LLGRVVLNPPALPFR 105

RESULT 10
ID_PLSX_CLOTE STANDARD; PRT; 334 AA.
AC Q95NO;
DT DT 10-OCT-2003 (Rel. 42, Created)
DT DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fatty acid/phospholipid synthesis protein plsx.
GN PLSX OR CTC01242.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1513;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=Massechusetts / E98;
RX MEDLINE=2245723; PubMed=12552129;
RA Rueggemann H., Baumer S., Fricke W.F., Wiesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease."
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -I- FUNCTION: Not known, probably involved in fatty acid or
CC phospholipid synthesis [By similarity].
CC -I- SIMILARITY: Belongs to the plsx family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
-----
CC EMBL; AB015940; AAC35810.1; ALT_INIT.
CC HAMAP; MF_00019; ; 1.
CC InterPro; IPR003664; FA_synthesis.
CC Pfam; PF02504; FA_synthesis; 1.
CC ProDom; PD006974; FA_synthesis; 1.
CC Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SQ SEQUENCE 334 AA; 36422 MW; 5382B255167FF004 CRC64;

Query Match 43.6%; Score 48; DB 1; Length 334;
Best Local Similarity 57.9%; Pred. No. 7;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 GSVIVIGRIISGRPAIVP 21
   ||:|||||:|:
Db 110 GSLFIVIRIKGIDRPAIP 128

RESULT 11
MAUD METME STANDARD; PRT; 211 AA.
ID ID MAUD METME
AC AC Q50232;
DT DT 01-NOV-1997 (Rel. 35, Created)
```

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Methylamine utilization protein maud.  
 GN MAUD.  
 OS Methylophilus methylotrophus (Bacterium W3A1).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales;  
 OC Methylophilaceae; Methylophilus.  
 CX NCBI\_TaxID=17;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94292427; PubMed=8021188;  
 RA CHRISTENSEN A.Y., McIntire W.S., Mathews F.S., Lidstrom M.E.;  
 RT "Organization of the methylamine utilization (mau) genes in  
 RT Methylophilus methylotrophus W3A1-NS."  
 RL J. Bacteriol. 176:4073-4080(1994).  
 CC -!- FUNCTION: MAY BE SPECIFICALLY INVOLVED IN THE PROCESSING,  
 CC TRANSPORT, AND/OR MATURATION OF THE WADH BETA-SUBUNIT.  
 CC -!- PATHWAY: Methylamine utilization.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L26407; AAB46950.1; -;  
 DR PIR; T10072.  
 DR InterPro; IPR006663; Thioresox\_dom2.  
 DR Transmembrane.  
 KW TRANSMEM  
 FT TRANSMEM 5 25 POTENTIAL.  
 SQ SEQUENCE 211 AA; 23083 MW; ABA7E16BE4CE4E58 CRC64;  
 Query Match 42.7%; Score 47; DB 1; Length 211;  
 Best Local Similarity 50.0%; Pred. No. 6.5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 GSVVIVGRILSGRPAIV 20  
 DB 66 GEPVLVGRSTPGPSLL 83  
 RESULT 12  
 ID -!-SPD COREF STANDARD; PRT; 248 AA.  
 AC QBPML3;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MCP)  
 DE cytidyltransferase) (MCT).  
 GN ISPD OR CE2521.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 CX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA MEDLINE=227273752; PubMed=12840036;  
 RA Nishio Y., Nakamura Y., Kawarayashi Y., Usuda Y., Kimura E.,  
 RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
 RA Gojobori T.;  
 RT "Comparative complete genome sequence analysis of the amino acid  
 RT replacements responsible for the thermostability of Corynebacterium  
 RT efficiens."  
 RL Genome Res. 13:1572-1579(2003).  
 CC -!- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-methyl-  
 CC D-erythritol from CTP and 2C-methyl-D-erythritol 4-phosphate (By  
 CC similarity).  
 CC -!- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =

CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -!- SIMILARITY: Belongs to the ispd family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AP005222; BAC19331.1; -;  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR TIGRFAMs; TIGR00453; ispd; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 KW Transferase; Nucleotidyltransferase; Isoprene biosynthesis;  
 KW Complete proteome.  
 KW SEQUENCE 248 AA; 26727 MW; AC66FF58948CEFF3 CRC64;  
 Query Match 41.8%; Score 46; DB 1; Length 248;  
 Best Local Similarity 36.8%; Pred. No. 11;  
 Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
 QY 3 GSVVIVGRILSGRPAIV 21  
 DB 122 GMIAVAVMADGAPAVIP 140  
 RESULT 13  
 ID -!-SPD COREF STANDARD; PRT; 332 AA.  
 AC QBR9V6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Fatty acid/phospholipid synthesis protein plsx.  
 GN PLXS OR TTE1476.  
 OS Thermoanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
 CX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MB4 / JCM 11007;  
 RX MEDLINE=21992816; PubMed=1197336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome."  
 RL Genome Res. 12:689-700(2002).  
 CC -!- FUNCTION: Not known, probably involved in fatty acid or  
 CC phospholipid synthesis (By similarity).  
 CC -!- SIMILARITY: Belongs to the plsx family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE013105; AAM24698.1; AJT\_INIT.  
 DR HAMAP; MF\_00019; -; 1.  
 DR InterPro; IPR003664; FA\_synthesis.  
 DR Pfam; PF02504; FA\_synthesis; 1.  
 DR ProDom; PD006974; FA\_synthesis; 1.  
 DR TIGRFAMs; TIGR00182; plsx; 1.  
 KW Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.  
 KW SEQUENCE 332 AA; 35989 MW; 0E71F6915DD1E9D8 CRC64;

RESULT 15  
556 CAMME  
- C756 CAMME STANDARD; PRT; 523 AA.  
O04772;  
15-DEC-1998 (Rel. 37, Created)  
15-DEC-1998 (Rel. 37, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Flavonoid 3', 5'-hydroxylase (EC 1.14.-.- (F3',5'H) (Cytochrome P450  
75A6).  
CYP75A6.  
Campanula medium (Canterbury bells)

Search completed: July 6, 2004, 09:28:37  
Job time : 12 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:28:15 ; Search time 29 Seconds  
(without alignments)  
250.239 Million cell updates/sec

Title: US-10-070-220-13  
Perfect score: 110  
Sequence: 1 KKGSVIVIGRIILSGRPAIVPRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 93    | 84.5        | 3010   | 12 Q99AU2 | Q99au2 hepatitis c |
| 2          | 92    | 83.6        | 138    | 12 Q68218 | Q68218 hepatitis c |
| 3          | 92    | 83.6        | 138    | 12 Q68244 | Q68244 hepatitis c |
| 4          | 92    | 83.6        | 138    | 12 Q68242 | Q68242 hepatitis c |
| 5          | 92    | 83.6        | 138    | 12 Q68216 | Q68216 hepatitis c |
| 6          | 92    | 83.6        | 138    | 12 Q68205 | Q68205 hepatitis c |
| 7          | 92    | 83.6        | 138    | 12 Q68217 | Q68217 hepatitis c |
| 8          | 92    | 83.6        | 172    | 12 Q68218 | Q68218 hepatitis c |
| 9          | 92    | 83.6        | 172    | 12 Q68217 | Q68217 hepatitis c |
| 10         | 92    | 83.6        | 172    | 12 Q68216 | Q68216 hepatitis c |
| 11         | 92    | 83.6        | 172    | 12 Q68215 | Q68215 hepatitis c |
| 12         | 92    | 83.6        | 172    | 12 Q68214 | Q68214 hepatitis c |
| 13         | 92    | 83.6        | 172    | 12 Q68213 | Q68213 hepatitis c |
| 14         | 92    | 83.6        | 271    | 12 Q68212 | Q68212 hepatitis c |
| 15         | 92    | 83.6        | 425    | 12 Q68344 | Q68344 hepatitis c |
| 16         | 92    | 83.6        | 1186   | 12 Q61755 | Q61755 hepatitis c |

|    |    |      |      |           |                     |
|----|----|------|------|-----------|---------------------|
| 17 | 92 | 83.6 | 2284 | 12 Q81817 | Q81817 hepatitis c  |
| 18 | 92 | 83.6 | 3010 | 12 Q9QIX1 | Q9qix1 hepatitis c  |
| 19 | 92 | 83.6 | 3010 | 12 Q9J3G1 | Q9j3g1 hepatitis c  |
| 20 | 92 | 83.6 | 3010 | 12 Q9J3I1 | Q9j3i1 hepatitis c  |
| 21 | 92 | 83.6 | 3010 | 12 Q9J3H1 | Q9j3h1 hepatitis c  |
| 22 | 92 | 83.6 | 3010 | 12 Q9DTE2 | Q9dte2 hepatitis c  |
| 23 | 92 | 83.6 | 3010 | 12 Q9J3G2 | Q9j3g2 hepatitis c  |
| 24 | 92 | 83.6 | 3010 | 12 Q9QIY7 | Q9qiyl7 hepatitis c |
| 25 | 92 | 83.6 | 3010 | 12 Q9QIX2 | Q9qix2 hepatitis c  |
| 26 | 92 | 83.6 | 3010 | 12 Q93016 | Q93016 hepatitis c  |
| 27 | 92 | 83.6 | 3010 | 12 Q68949 | Q68949 hepatitis c  |
| 28 | 92 | 83.6 | 3010 | 12 Q81989 | Q81989 hepatitis c  |
| 29 | 92 | 83.6 | 3010 | 12 Q9J3I0 | Q9j3i0 hepatitis c  |
| 30 | 92 | 83.6 | 3010 | 12 Q9QIY8 | Q9qiyl8 hepatitis c |
| 31 | 92 | 83.6 | 3010 | 12 P89366 | P89366 hepatitis c  |
| 32 | 92 | 83.6 | 3010 | 12 Q9QIX8 | Q9qix8 hepatitis c  |
| 33 | 92 | 83.6 | 3010 | 12 Q9DTE8 | Q9dte8 hepatitis c  |
| 34 | 92 | 83.6 | 3010 | 12 Q9DTE1 | Q9dte1 hepatitis c  |
| 35 | 92 | 83.6 | 3010 | 12 Q9J3G7 | Q9j3g7 hepatitis c  |
| 36 | 92 | 83.6 | 3010 | 12 Q9QIX7 | Q9qiyl7 hepatitis c |
| 37 | 92 | 83.6 | 3010 | 12 Q81541 | Q81541 hepatitis c  |
| 38 | 92 | 83.6 | 3010 | 12 Q9J3H6 | Q9j3h6 hepatitis c  |
| 39 | 92 | 83.6 | 3010 | 12 Q9J3G0 | Q9j3g0 hepatitis c  |
| 40 | 92 | 83.6 | 3011 | 12 Q9DTE8 | Q9dte8 hepatitis c  |
| 41 | 92 | 83.6 | 3011 | 12 Q9DTE3 | Q9dte3 hepatitis c  |
| 42 | 92 | 83.6 | 3013 | 12 Q9QNC0 | Q9qnc0 hepatitis c  |
| 43 | 92 | 83.6 | 3013 | 12 Q9QIY0 | Q9qiyl0 hepatitis c |
| 44 | 92 | 83.6 | 3014 | 12 Q86614 | Q86614 hepatitis c  |
| 45 | 92 | 83.6 | 3015 | 12 Q9WPH5 | Q9wph5 hepatitis c  |

#### ALIGNMENTS

#### RESULT 1

Q99AU2 ID Q99AU2 PRELIMINARY; PRT; 3010 AA.  
AC Q99AU2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polypeptide.  
OS Hepatitis C virus type 1b.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31647;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SFRANK-chimera of HCV-BK;  
RA Thomson M., Nascimben M., Gonzales S., Murthy K., Rehmann B.,  
RA Liang J.;  
RT "Analyses of viral sequences and virus-specific immune responses  
RT during serial passage of an infectious hepatitis C virus serotype 1b  
RT clone in chimpanzees."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA (BY SIMILARITY).  
CC EMBL; AF333324; AAK08509.1; -.  
CC PIR; A61196; A61196.  
CC PIR; PQ0246; PQ0246.  
CC PIR; PQ0804; PQ0804.  
CC PIR; PS0329; PS0329.  
CC HSSP; P26663; INS3.  
CC DR GO; GO:0016021; C:integral to membrane; IEA.  
CC DR GO; GO:0019028; C:viral capsid; IEA.  
CC DR GO; GO:0019031; C:viral envelope; IEA.  
CC DR GO; GO:0005524; F:ATP binding; IEA.  
CC DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
CC DR GO; GO:0005489; F:electron transporter activity; IEA.  
CC DR GO; GO:0003723; F:RNA binding; IEA.  
CC DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.



```

RA  Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT  "Prevalence of hepatitis C virus sequence variants in South-East
RL  Asia.";
RL  J. Gen. Virol. 76:211-215(1995).
DR  EMBL; U14278; AAC53967.1; -.
DR  HSP; P27958; IHEI.
DR  InterPro; IPR000745; HCV_NS4a.
DR  InterPro; IPR001490; HCV_NS4b.
DR  Pfam; PF01006; HCV_NS4a; 1.
DR  Pfam; PF01001; HCV_NS4b; 1.
FT  NON_TER 1
FT  NON_TER 138
SQ  SEQUENCE 138 AA; 15281 MW; CD5B5B3834C6070D CRC64;

Query Match      83.6%; Score 92; DB 12; Length 138;
Best Local Similarity 90.5%; Pred. No. 4.6e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 GSVVIVGRILLSGRPAIVPRR 23
DB  52 GSVVIVGRILLSGRPAIVPR 72

RESULT 5
Q68216
ID  Q68216 PRELIMINARY; PRT; 138 AA.
AC  Q68216;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Nonstructural protein (Fragment).
GN  NS4.
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=11103;
RN  [1]_TaxID=11103;
RP  SEQUENCE FROM N.A.
RC  STRAIN=1b;
RX  MEDLINE=95146953; PubMed=7844535;
RA  Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT  "Prevalence of hepatitis C virus sequence variants in South-East
RL  Asia.";
RL  J. Gen. Virol. 76:211-215(1995).
DR  EMBL; U14252; AAC53941.1; -.
DR  HSP; P27958; IHEI.
DR  InterPro; IPR000745; HCV_NS4a.
DR  InterPro; IPR001490; HCV_NS4b.
DR  Pfam; PF01006; HCV_NS4a; 1.
DR  Pfam; PF01001; HCV_NS4b; 1.
FT  NON_TER 1
FT  NON_TER 138
SQ  SEQUENCE 138 AA; 15081 MW; 7ED533A7D169FBLA CRC64;

Query Match      83.6%; Score 92; DB 12; Length 138;
Best Local Similarity 90.5%; Pred. No. 4.6e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 GSVVIVGRILLSGRPAIVPRR 23
DB  52 GSVVIVGRILLSGRPAIVPR 72

RESULT 6
Q68205
ID  Q68205 PRELIMINARY; PRT; 138 AA.
AC  Q68205;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Nonstructural protein (Fragment).
GN  NS4.
OS  Hepatitis C virus.

```

```

OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=11103;
RN  [1]_TaxID=11103;
RP  SEQUENCE FROM N.A.
RC  STRAIN=1b;
RX  MEDLINE=95146953; PubMed=7844535;
RA  Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT  "Prevalence of hepatitis C virus sequence variants in South-East
RL  Asia.";
RL  J. Gen. Virol. 76:211-215(1995).
DR  EMBL; U14241; AAC53930.1; -.
DR  HSP; P27958; IHEI.
DR  InterPro; IPR000745; HCV_NS4a.
DR  InterPro; IPR001490; HCV_NS4b.
DR  Pfam; PF01006; HCV_NS4a; 1.
DR  Pfam; PF01001; HCV_NS4b; 1.
FT  NON_TER 1
FT  NON_TER 138
SQ  SEQUENCE 138 AA; 15205 MW; 6D376B2DD86EADAA CRC64;

Query Match      83.6%; Score 92; DB 12; Length 138;
Best Local Similarity 90.5%; Pred. No. 4.6e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 GSVVIVGRILLSGRPAIVPRR 23
DB  52 GSVVIVGRILLSGRPAIVPR 72

RESULT 7
Q68217
ID  Q68217 PRELIMINARY; PRT; 138 AA.
AC  Q68217;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Nonstructural protein (Fragment).
GN  NS4.
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=11103;
RN  [1]_TaxID=11103;
RP  SEQUENCE FROM N.A.
RC  STRAIN=1b;
RX  MEDLINE=95146953; PubMed=7844535;
RA  Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT  "Prevalence of hepatitis C virus sequence variants in South-East
RL  Asia.";
RL  J. Gen. Virol. 76:211-215(1995).
DR  EMBL; U14253; AAC53942.1; -.
DR  HSP; P27958; IHEI.
DR  InterPro; IPR000745; HCV_NS4a.
DR  InterPro; IPR001490; HCV_NS4b.
DR  Pfam; PF01006; HCV_NS4a; 1.
DR  Pfam; PF01001; HCV_NS4b; 1.
FT  NON_TER 1
FT  NON_TER 138
SQ  SEQUENCE 138 AA; 15130 MW; 2AF1E92DDC7B741D CRC64;

Query Match      83.6%; Score 92; DB 12; Length 138;
Best Local Similarity 90.5%; Pred. No. 4.6e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 GSVVIVGRILLSGRPAIVPRR 23
DB  52 GSVVIVGRILLSGRPAIVPR 72

RESULT 8
Q81582
ID  Q81582 PRELIMINARY; PRT; 172 AA.

```

```
AC Q81582;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD4;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86776; AAA45625.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
SQ SEQUENCE 172 AA; 18941 MW; 08DC37AE66CD84D6 CRC64;

Query Match 83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23
Db 107 GSVVIVGRILSGRPAIVPR 127

RESULT 9
Q81584
ID Q81584 PRELIMINARY; PRT; 172 AA.
AC Q81584;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD5;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86778; AAA45627.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
SQ SEQUENCE 172 AA; 18941 MW; 08DC37AE66CD84D6 CRC64;

Query Match 83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23
Db 107 GSVVIVGRILSGRPAIVPR 127

RESULT 10
Q81574
ID Q81574 PRELIMINARY; PRT; 172 AA.
AC Q81574;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD3;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86767; AAA45617.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
SQ SEQUENCE 172 AA; 18941 MW; 08DC37AE66CD84D6 CRC64;

Query Match 83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23
Db 107 GSVVIVGRILSGRPAIVPR 127

RESULT 11
Q81578
ID Q81578 PRELIMINARY; PRT; 172 AA.
AC Q81578;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD4-2;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86772; AAA45621.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
SQ SEQUENCE 172 AA; 18941 MW; 08DC37AE66CD84D6 CRC64;

Query Match 83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23
Db 107 GSVVIVGRILSGRPAIVPR 127
```

```

Db      107 GSVVIVGRILSGRPAIVPR 127
|||||
RESULT 12
Q81583 PRELIMINARY; PRT; 172 AA.
ID Q81583
AC Q81583;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD6-1;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86777; AAA45624.1; -.
DR HSSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 172
SQ SEQUENCE 172 AA; 18943 MW; 49D8356DC338179E CRC64;

Query Match      83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPR 23
Db      107 GSVVIVGRILSGRPAIVPR 127
|||||
RESULT 13
Q81581 PRELIMINARY; PRT; 172 AA.
ID Q81581
AC Q81581;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD6-2;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86775; AAA45624.1; -.
DR HSSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 172
SQ SEQUENCE 172 AA; 18941 MW; 08DC37AE66CD84D6 CRC64;

Query Match      83.6%; Score 92; DB 12; Length 172;
Best Local Similarity 90.5%; Pred. No. 5.8e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPR 23
Db      107 GSVVIVGRILSGRPAIVPR 127
|||||
RESULT 14
Q81573 PRELIMINARY; PRT; 271 AA.
ID Q81573
AC Q81573;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS3-NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD2;
RA Miller H.M., Goesser T., Pfaff E., Theilmann L.;
RT "Characterization of German hepatitis C virus isolates: Comparison
RT with different isolate, genetic variability of HCV within a country
RT and even in a single patient and antigenicity of expressed core
RT protein.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86766; AAA45616.1; -.
DR HSSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
FT NON_TER 1
FT NON_TER 271
SQ SEQUENCE 271 AA; 29796 MW; 72EBEC54E6877CD4 CRC64;

Query Match      83.6%; Score 92; DB 12; Length 271;
Best Local Similarity 90.5%; Pred. No. 9.1e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPR 23
Db      151 GSVVIVGRILSGRPAIVPR 171
|||||
RESULT 15
Q68344 PRELIMINARY; PRT; 425 AA.
ID Q68344
AC Q68344;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-K1;
RA Paik S.H., Yang J.M.;
RT "Hepatitis C virus genome, complete NS4 and part of NS3 and NS5
RT sequence.";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U26687; AAA79971.1; -.
DR HSSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR Pfam; PF01006; HCV_NS4a; 1.

```

Tue Jul 6 10:45:10 2004

DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
FT NON\_TER 1 1 NS3 PROTEIN.  
FT CHAIN 1 >9 NS4 PROTEIN.  
FT CHAIN 10 >406 NS5 PROTEIN.  
FT CHAIN 407 >425  
FT NON\_TER 425 425  
SQ SEQUENCE 425 AA; 45366 MW; 1A2474932E0EB262 CRC64;  
Query Match 83.6%; Score 92; DB 12; Length 425;  
Best Local Similarity 90.5%; Pred. No. 1.4e-05;  
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 3 GSVVIVGRIILSGRPAIVPRR 23  
Db 72 GSVVIVGRIILSGRPAIVPRR 92

Search completed: July 6, 2004, 09:29:39  
Job time : 32 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

\* Run on: July 6, 2004, 09:28:15 ; Search time 15 Seconds  
(without alignments)  
79.160 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110

Sequence: 1 KXGSVVIVGRILSGRPAIVPRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*

5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*

6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 99    | 90.0        | 23     | 3     | US-08-731-336-1   |
| 2          | 99    | 90.0        | 23     | 4     | US-09-257-667-1   |
| 3          | 99    | 90.0        | 23     | 4     | US-09-881-239-8   |
| 4          | 99    | 90.0        | 23     | 4     | US-09-881-654-7   |
| 5          | 93    | 84.5        | 34     | 1     | US-08-700-356-3   |
| 6          | 93    | 84.5        | 34     | 2     | US-08-936-865-3   |
| 7          | 93    | 84.5        | 54     | 1     | US-08-700-356-2   |
| 8          | 93    | 84.5        | 54     | 2     | US-08-936-865-2   |
| 9          | 93    | 84.5        | 54     | 3     | US-09-198-723A-24 |
| 10         | 93    | 84.5        | 54     | 4     | US-09-684-881-24  |
| 11         | 93    | 84.5        | 247    | 1     | US-08-324-977-44  |
| 12         | 93    | 84.5        | 247    | 2     | US-08-384-616-44  |
| 13         | 93    | 84.5        | 247    | 2     | US-08-904-686A-44 |
| 14         | 93    | 84.5        | 247    | 3     | US-09-315-850-44  |
| 15         | 93    | 84.5        | 1692   | 3     | US-09-263-933-4   |
| 16         | 93    | 84.5        | 1692   | 3     | US-09-263-933-11  |
| 17         | 93    | 84.5        | 1692   | 3     | US-09-263-933-18  |
| 18         | 93    | 84.5        | 1692   | 4     | US-09-919-901-4   |
| 19         | 93    | 84.5        | 1692   | 4     | US-09-919-901-11  |
| 20         | 93    | 84.5        | 1692   | 4     | US-09-919-901-18  |
| 21         | 93    | 84.5        | 2013   | 1     | US-08-324-977-12  |
| 22         | 93    | 84.5        | 2013   | 2     | US-08-384-616-12  |
| 23         | 93    | 84.5        | 2013   | 2     | US-08-904-686A-12 |
| 24         | 93    | 84.5        | 2201   | 3     | US-09-315-850-12  |
| 25         | 93    | 84.5        | 2201   | 4     | US-08-952-981A-2  |
| 26         | 93    | 84.5        | 2307   | 3     | US-09-263-933-2   |
| 27         | 93    | 84.5        | 2307   | 3     | US-09-263-933-9   |

|    |    |      |      |   |                   |
|----|----|------|------|---|-------------------|
| 28 | 93 | 84.5 | 2307 | 3 | US-09-263-933-16  |
| 29 | 93 | 84.5 | 2307 | 4 | US-09-919-901-2   |
| 30 | 93 | 84.5 | 2307 | 4 | US-09-919-901-9   |
| 31 | 93 | 84.5 | 2307 | 4 | US-09-919-901-16  |
| 32 | 93 | 84.5 | 2620 | 1 | US-08-324-977-32  |
| 33 | 93 | 84.5 | 2620 | 2 | US-08-384-616-32  |
| 34 | 93 | 84.5 | 2620 | 2 | US-08-904-686A-32 |
| 35 | 93 | 84.5 | 2620 | 3 | US-09-315-850-32  |
| 36 | 93 | 84.5 | 2621 | 1 | US-08-324-977-36  |
| 37 | 93 | 84.5 | 2621 | 2 | US-08-384-616-36  |
| 38 | 93 | 84.5 | 2621 | 2 | US-08-904-686A-36 |
| 39 | 93 | 84.5 | 2621 | 3 | US-09-315-850-36  |
| 40 | 93 | 84.5 | 3010 | 1 | US-08-324-977-2   |
| 41 | 93 | 84.5 | 3010 | 1 | US-08-324-977-14  |
| 42 | 93 | 84.5 | 3010 | 2 | US-08-384-616-2   |
| 43 | 93 | 84.5 | 3010 | 2 | US-08-384-616-14  |
| 44 | 93 | 84.5 | 3010 | 2 | US-08-904-686A-2  |
| 45 | 93 | 84.5 | 3010 | 2 | US-08-904-686A-14 |

#### ALIGNMENTS

RESULT 1  
US-08-731-336-1  
; Sequence 1, Application US/08731336  
; Patent No. 6153579  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Joseph L.  
; APPLICANT: Morgenstern, Kurt A.  
; APPLICANT: Lin, Chao  
; APPLICANT: Fox, Ted  
; APPLICANT: Thomson, John A.  
; TITLE OF INVENTION: CRYSTALLIZABLE COMPOSITIONS COMPRISING A  
; HEPATITIS C VIRUS NS3 PROTEASE DOMAIN/NS4A COMPLEX  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/731,336  
; FILING DATE: 18-OCT-1996  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: VPI96-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; US-08-731-336-1

Query Match 90.0%; Score 99; DB 3; Length 23;  
Best Local Similarity 78.3%; Pred. No. 2.1e-09;





```
/
/ GENERAL INFORMATION:
/ APPLICANT: DE FRANCESCO, Raffaele
/ APPLICANT: FAILLA, Cristina
/ APPLICANT: TOMEI, Licia
/ TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE
/ TITLE OF INVENTION: PROTEOLYTIC ACTIVITY OF THE NS3 HEPATITIS C VIRUS (HCV)
/ NUMBER OF SEQUENCES: 3
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
/ STREET: 419 Seventh Street, N.W., Suite 300
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/700,356
/ FILING DATE: 23-AUG-1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BROWDY, Roger L.
/ REGISTRATION NUMBER: 25,618
/ REFERENCE/DOCKET NUMBER: DE FRANCESCO-1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 34 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ FEATURE:
/ OTHER INFORMATION: /note= "Xaa at position 34 means
/ OTHER INFORMATION: Abu (2-Aminobutyric acid)"
/
/ US-08-700-356-3
/
/ Query Match 84.5%; Score 93; DB 1; Length 34;
/ Best Local Similarity 95.2%; Pred. No. 3e-08;
/ Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
/
/ QY 3 GSVVIVGRILLSGRPAIVPRR 23
/ Db 1 GSVVIVGRILLSGRPAIVPDR 21
/
/ RESULT 6
/ US-08-936-865-3
/ Sequence 3, Application US/0893685
/ Patent No. 5861297
/ GENERAL INFORMATION:
/ APPLICANT: Sardana, Vinod V
/ APPLICANT: Blue, Jeffrey T
/ TITLE OF INVENTION: DETERGENT-FREE HEPATITIS C PROTEASE
/ NUMBER OF SEQUENCES: 3
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MERCK & CO., INC.
/ STREET: P.O. Box 2000, 126 E. Lincoln Ave.
/ CITY: Rahway
/ STATE: NJ
/ COUNTRY: US
/ ZIP: 07065-0907
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/
/ APPLICATION NUMBER: US/08/936,865
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BROWDY, Roger L.
/ REGISTRATION NUMBER: 25,618
/ REFERENCE/DOCKET NUMBER: DE FRANCESCO=1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 54 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/
/ APPLICATION NUMBER: US/08/936,865
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Aylor, Sylvia A
/ REGISTRATION NUMBER: 36,436
/ REFERENCE/DOCKET NUMBER: 19691
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 908-594-4909
/ TELEFAX: 908-594-4720
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 34 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ IMMEDIATE SOURCE:
/ LIBRARY: Cofactor of NS3 serine protease
/ CLONE: Solid phase peptide synthesis
/
/ US-08-936-865-3
/
/ Query Match 84.5%; Score 93; DB 2; Length 34;
/ Best Local Similarity 95.2%; Pred. No. 3e-08;
/ Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
/
/ QY 3 GSVVIVGRILLSGRPAIVPRR 23
/ Db 1 GSVVIVGRILLSGRPAIVPDR 21
/
/ RESULT 7
/ US-08-700-356-2
/ Sequence 2, Application US/08700356
/ Patent No. 5739002
/ GENERAL INFORMATION:
/ APPLICANT: DE FRANCESCO, Raffaele
/ APPLICANT: FAILLA, Cristina
/ APPLICANT: TOMEI, Licia
/ TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE
/ TITLE OF INVENTION: PROTEOLYTIC ACTIVITY OF THE NS3 HEPATITIS C VIRUS (HCV)
/ NUMBER OF SEQUENCES: 3
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
/ STREET: 419 Seventh Street, N.W., Suite 300
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/700,356
/ FILING DATE: 23-AUG-1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BROWDY, Roger L.
/ REGISTRATION NUMBER: 25,618
/ REFERENCE/DOCKET NUMBER: DE FRANCESCO=1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 54 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-700-356-2

Query Match      84.5%; Score 93; DB 1; Length 54;
Best Local Similarity 95.2%; Pred. No. 5e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPRR 23
Db      21 GSVVIVGRILSGRPAIVPDR 41

RESULT 8
US-08-936-865-2
; Sequence 2, Application US/08936865
; Patent No. 5861297
; GENERAL INFORMATION:
; APPLICANT: Sardana, Vinod V
; APPLICANT: Blue, Jeffrey T
; TITLE OF INVENTION: DETERGENT-FREE HEPATITIS C PROTEASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCK & CO., INC.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,865
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ayler, Sylvia A
; REGISTRATION NUMBER: 36,436
; REFERENCE/DOCKET NUMBER: 19691
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-4909
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; IMMEDIATE SOURCE:
; LIBRARY: CDNA clone (See Seq. ID No. 58612971)
; CLONE: NS4A Protein
US-08-936-865-2

Query Match      84.5%; Score 93; DB 2; Length 54;
Best Local Similarity 95.2%; Pred. No. 5e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPRR 23
Db      21 GSVVIVGRILSGRPAIVPDR 41

RESULT 9
US-09-198-723A-24
; Sequence 24, Application US/09198723A
; Patent No. 6211338
```

```
; GENERAL INFORMATION:
; APPLICANT: Malcolml, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JE0800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-198-723A-24

Query Match      84.5%; Score 93; DB 3; Length 54;
Best Local Similarity 95.2%; Pred. No. 5e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GSVVIVGRILSGRPAIVPRR 23
Db      21 GSVVIVGRILSGRPAIVPDR 41

RESULT 10
US-09-684-881-24
; Sequence 24, Application US/09684881
; Patent No. 6653127
; GENERAL INFORMATION:
; APPLICANT: Malcolml, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/684,881
; FILING DATE: 06-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723
; FILING DATE: 24 NOV 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-684-881-24

Query Match      84.5%; Score 93; DB 4; Length 54;
Best Local Similarity 95.2%; Pred. NO. 5e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILLSGRPAIVPRR 23
Db 21 GSVVIVGRILLSGRPAIVPDR 41

RESULT 11
US-08-324-977-44
; Sequence 44, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLeland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; APPLICATION NUMBER: US 08/099,706

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/684,881
; FILING DATE: 06-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723
; FILING DATE: 24 NOV 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-684-881-24

Query Match      84.5%; Score 93; DB 4; Length 54;
Best Local Similarity 95.2%; Pred. NO. 5e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILLSGRPAIVPRR 23
Db 21 GSVVIVGRILLSGRPAIVPDR 41

RESULT 12
US-08-384-616-44
; Sequence 44, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLeland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,616
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; APPLICATION NUMBER: US 08/099,706
```

; PRIOR APPLICATION DATA: US 07/635,451  
; APPLICATION NUMBER: US 07/635,451  
; FILING DATE: 28-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stevens-Smith, Theresa M.  
; REGISTRATION NUMBER: 36,281  
; REFERENCE/DOCKET NUMBER: 900703B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-2930  
; TELEFAX: (202) 887-0357  
; TELEX: 440142  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-384-616-44

Query Match 84.5%; Score 93; DB 2; Length 247;  
Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
|||||  
DB 63 GSVVIVGRILSGRPAIVPDR 83

## RESULT 13

US-08-904-686A-44  
; Sequence 44, Application US/08904686A  
; Patent No. 5998130  
; GENERAL INFORMATION:  
; APPLICANT: OKAYAMA, Hiroto  
; APPLICANT: FUKU, Isao  
; APPLICANT: MORI, Chisato  
; APPLICANT: TAKAMIZAWA, Akahisa  
; APPLICANT: YOSHIDA, Iwao  
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLand &  
; ADDRESSEE: Naughton  
; STREET: 1725 K St. N.W. Suite 1000  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/904,686A  
; FILING DATE: 01-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/324,977  
; FILING DATE: 18-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-167466  
; FILING DATE: 25-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-230921  
; FILING DATE: 31-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-305605  
; FILING DATE: 09-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/099,706  
; FILING DATE: 30-JUL-1993  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/769,996  
; FILING DATE: 02-OCT-1991  
; PRIOR APPLICATION DATA: US 07/635,451  
; APPLICATION NUMBER: US 07/635,451  
; FILING DATE: 28-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McLeLand, Le-Nhung  
; REGISTRATION NUMBER: 31,541  
; REFERENCE/DOCKET NUMBER: 900703G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-2930  
; TELEFAX: (202) 887-0357  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-904-686A-44

Query Match 84.5%; Score 93; DB 2; Length 247;  
Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
|||||  
DB 63 GSVVIVGRILSGRPAIVPDR 83

## RESULT 14

US-09-315-850-44  
; Sequence 44, Application US/09315850  
; Patent No. 6217872  
; GENERAL INFORMATION:  
; APPLICANT: OKAYAMA, Hiroto  
; APPLICANT: FUKU, Isao  
; APPLICANT: MORI, Chisato  
; APPLICANT: TAKAMIZAWA, Akahisa  
; APPLICANT: YOSHIDA, Iwao  
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLand &  
; ADDRESSEE: Naughton  
; STREET: 1725 K St. N.W. Suite 1000  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/315,850  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/904,686  
; FILING DATE: 01-AUG-1997  
; APPLICATION NUMBER: US 08/324,977  
; FILING DATE: 18-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-167466  
; FILING DATE: 25-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-230921  
; FILING DATE: 31-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-305605  
; FILING DATE: 09-NOV-1990  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/099,706  
FILING DATE: 30-JUL-1993  
PRIOR APPLICATION DATA: US 07/769,996  
FILING DATE: 02-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/635,451  
FILING DATE: 28-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Mclelland, Le-Nhung  
REGISTRATION NUMBER: 31,541  
REFERENCE/DOCKET NUMBER: 900703G  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 247 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-315-850-44

Query Match 84.5%; Score 93; DB 3; Length 247;  
Best Local Similarity 95.2%; Pred. No. 2.7e-07;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
|||||  
Db 63 GSVVIVGRILSGRPAIVPDR 83

## RESULT 15

US-09-263-933-4  
; Sequence 4, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263,933  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 09/129,611  
; EARLIER FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
US-09-263-933-4

Query Match 84.5%; Score 93; DB 3; Length 1692;  
Best Local Similarity 95.2%; Pred. No. 2.3e-06;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
|||||  
Db 957 GSVVIVGRILSGRPAIVPDR 977

Search completed: July 6, 2004, 09:30:01  
Job time : 15 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 09:29:43 ; Search time 301 Seconds  
(without alignments)  
23.786 Million cell updates/sec

Title: US-10-070-220-13

Perfect score: 110  
Sequence: 1 KKGSVIVGRILISGRPAIVPRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description      |
|------------|-------|-------------|--------|----|------------------|
| 1          | 99    | 90.0        | 23     | 9  | US-09-881-654-7  |
| 2          | 99    | 90.0        | 23     | 9  | US-09-881-239-8  |
| 3          | 99    | 90.0        | 23     | 12 | US-10-637-323-7  |
| 4          | 99    | 90.0        | 23     | 12 | US-09-728-653-6  |
| 5          | 99    | 90.0        | 23     | 14 | US-10-010-184A-7 |
| 6          | 99    | 90.0        | 23     | 14 | US-10-015-328-6  |
| 7          | 99    | 90.0        | 23     | 16 | US-10-643-853-8  |
| 8          | 93    | 84.5        | 1692   | 10 | US-09-919-901-4  |
| 9          | 93    | 84.5        | 1692   | 10 | US-09-919-901-11 |
| 10         | 93    | 84.5        | 1692   | 10 | US-09-919-901-18 |
| 11         | 93    | 84.5        | 1692   | 14 | US-10-191-966-4  |
| 12         | 93    | 84.5        | 1692   | 14 | US-10-191-966-11 |
| 13         | 93    | 84.5        | 1692   | 14 | US-10-191-966-18 |
| 14         | 93    | 84.5        | 2201   | 13 | US-10-085-476-2  |
| 15         | 93    | 84.5        | 2307   | 10 | US-09-919-901-2  |

|    |    |      |      |    |                   |                    |
|----|----|------|------|----|-------------------|--------------------|
| 16 | 93 | 84.5 | 2307 | 10 | US-09-919-901-9   | Sequence 9, Appli  |
| 17 | 93 | 84.5 | 2307 | 10 | US-09-919-901-16  | Sequence 16, Appli |
| 18 | 93 | 84.5 | 2307 | 14 | US-10-191-966-2   | Sequence 2, Appli  |
| 19 | 93 | 84.5 | 2307 | 14 | US-10-191-966-9   | Sequence 9, Appli  |
| 20 | 93 | 84.5 | 2307 | 14 | US-10-191-966-16  | Sequence 16, Appli |
| 21 | 92 | 83.6 | 1736 | 14 | US-10-328-127-2   | Sequence 2, Appli  |
| 22 | 92 | 83.6 | 1736 | 14 | US-10-328-206-2   | Sequence 2, Appli  |
| 23 | 92 | 83.6 | 1736 | 14 | US-10-328-206-2   | Sequence 2, Appli  |
| 24 | 89 | 80.9 | 1985 | 14 | US-10-259-275-40  | Sequence 40, Appli |
| 25 | 89 | 80.9 | 1985 | 14 | US-10-259-275-42  | Sequence 42, Appli |
| 26 | 89 | 80.9 | 3010 | 12 | US-10-639-150-2   | Sequence 2, Appli  |
| 27 | 86 | 78.2 | 2201 | 13 | US-10-467-000-1   | Sequence 1, Appli  |
| 28 | 86 | 78.2 | 2201 | 13 | US-10-029-907-3   | Sequence 3, Appli  |
| 29 | 85 | 77.3 | 79   | 9  | US-10-309-561-3   | Sequence 3, Appli  |
| 30 | 85 | 77.3 | 128  | 10 | US-09-758-308-3   | Sequence 3, Appli  |
| 31 | 85 | 77.3 | 128  | 10 | US-09-899-046-62  | Sequence 62, Appli |
| 32 | 85 | 77.3 | 128  | 10 | US-09-878-281-62  | Sequence 62, Appli |
| 33 | 85 | 77.3 | 481  | 10 | US-09-873-224-62  | Sequence 62, Appli |
| 34 | 85 | 77.3 | 481  | 10 | US-09-899-046-270 | Sequence 270, App  |
| 35 | 85 | 77.3 | 481  | 10 | US-09-878-281-270 | Sequence 270, App  |
| 36 | 85 | 77.3 | 481  | 10 | US-09-873-224-270 | Sequence 270, App  |
| 37 | 85 | 77.3 | 484  | 10 | US-09-899-046-198 | Sequence 198, App  |
| 38 | 85 | 77.3 | 484  | 10 | US-09-899-046-200 | Sequence 200, App  |
| 39 | 85 | 77.3 | 484  | 10 | US-09-878-281-198 | Sequence 198, App  |
| 40 | 85 | 77.3 | 484  | 10 | US-09-878-281-200 | Sequence 200, App  |
| 41 | 85 | 77.3 | 484  | 12 | US-09-873-224-198 | Sequence 198, App  |
| 42 | 83 | 75.5 | 51   | 9  | US-09-873-224-200 | Sequence 200, App  |
| 43 | 83 | 75.5 | 54   | 9  | US-09-921-397-91  | Sequence 91, Appli |
| 44 | 83 | 75.5 | 54   | 9  | US-09-929-955-7   | Sequence 7, Appli  |
| 45 | 83 | 75.5 | 54   | 9  | US-09-929-955-30  | Sequence 30, Appli |
|    |    |      | 54   | 12 | US-09-930-591-13  | Sequence 13, Appli |

## ALIGNMENTS

## RESULT 1

US-09-881-654-7  
; Sequence 7, Application US/09881654  
; Patent No. US20020146685A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARCANDEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMENTO, Carlos  
; APPLICANT: COURT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES  
; FILE REFERENCE: 2302-17039 / PPI7039.002  
; CURRENT APPLICATION NUMBER: US/09/881,654  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 60/212,082  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/280,811  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/280,867  
; PRIOR FILING DATE: 2001-04-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-09-881-654-7

Query Match 90.0%; Score 99; DB 9; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKGSVIVGRILISGRPAIVPRR 23  
Db 1 KKGSVIVGRIVLSGRPAIIPKK 23

```

Query Match      90.0%; Score 99; DB 9; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKGSVVIVGRILLSGRPAIVPRR 23
Db      1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 2
US-09-881-239-8
; Sequence 8, Application US/09881239
; Publication No. US20020192639A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIENTE, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
; FILE REFERENCE: 2302-16073 / PP16073.003
; CURRENT APPLICATION NUMBER: US/09/881,239
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide
US-09-881-239-8

Query Match      90.0%; Score 99; DB 9; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKGSVVIVGRILLSGRPAIVPRR 23
Db      1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 3
US-10-637-323-7
; Sequence 7, Application US/10637323
; Publication No. US20040063092A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIENTE, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PP17039.002
; CURRENT APPLICATION NUMBER: US/10/637,323
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: US/09/881,654
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/280,867
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide
US-10-637-323-7

Query Match      90.0%; Score 99; DB 12; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKGSVVIVGRILLSGRPAIVPRR 23
Db      1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 4
US-09-728-653-6
; Sequence 6, Application US/09728653
; Publication No. US20020123468A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Han, Amy Qi
; APPLICANT: Glunz, Peter W.
; TITLE OF INVENTION: Alpha-Ketoamide Inhibitors of Hepatitis C Virus NS3 Protease
; FILE REFERENCE: PH-7118
; CURRENT APPLICATION NUMBER: US/09/728,653
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/168,998
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A pept
; OTHER INFORMATION: ide synthesizer using readily available materials well known to o
; OTHER INFORMATION: rdinarily skilled artisans
US-09-728-653-6

Query Match      90.0%; Score 99; DB 12; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKGSVVIVGRILLSGRPAIVPRR 23
Db      1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 5
US-10-010-184A-7
; Sequence 7, Application US/10010184A
; Publication No. US20030008828A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Pharma Company
; APPLICANT: Priestly, et al.
; TITLE OF INVENTION: No. US20030008828A1el Lactam Inhibitors of Hepatitis C Virus NS3
; FILE REFERENCE: PH-7087-A
; CURRENT APPLICATION NUMBER: US/10/010,184A
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 09/626,286
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A pept
; OTHER INFORMATION: ide synthesizer using readily available materials well known to o
; OTHER INFORMATION: rdinarily skilled artisans
US-10-010-184A-7

Query Match      90.0%; Score 99; DB 14; Length 23;
Best Local Similarity 78.3%; Pred. No. 1.6e-08;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKGSVVIVGRILLSGRPAIVPRR 23
Db      1 KKGSVVIVGRIVLSGKPAIIPKK 23

```



RESULT 6  
US-10-015-328-6  
; Sequence 6, Application US/10015328  
; Publication No. US20030100768A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; APPLICANT: Han, Amy Qi  
; APPLICANT: Glunz, Peter W.  
; TITLE OF INVENTION: Imidazolidinones and Their Related Derivatives as Hepatitis C Virus Inhibitors  
; FILE REFERENCE: PH-7203  
; CURRENT APPLICATION NUMBER: US/10/015,328  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255,168  
; PRIOR FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: The synthesis of this peptide may be performed on an ABI 43A peptide synthesizer using readily available materials well known to those skilled in the art.  
; OTHER INFORMATION: rdinarily skilled artisans  
US-10-015-328-6

Query Match 90.0%; Score 99; DB 14; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
DB 1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 7  
US-10-643-853-8  
; Sequence 8, Application US/10643853  
; Publication No. US20040096822A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIEMENTO, Carlos  
; APPLICANT: COIT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY  
; FILE REFERENCE: 2302-16073 / PP16073.003  
; CURRENT APPLICATION NUMBER: US/10/643,853  
; CURRENT FILING DATE: 2003-08-19  
; PRIOR APPLICATION NUMBER: US/09/881,239  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: NS4A peptide  
US-10-643-853-8

Query Match 90.0%; Score 99; DB 16; Length 23;  
Best Local Similarity 78.3%; Pred. No. 1.6e-08;  
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGSVVIVGRILSGRPAIVPRR 23  
DB 1 KKGSVVIVGRIVLSGKPAIIPKK 23

RESULT 8  
US-09-919-901-4  
; Sequence 4, Application US/09919901  
; Publication No. US20030082518A1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263,933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: :  
US-09-919-901-4

Query Match 84.5%; Score 93; DB 10; Length 1692;  
Best Local Similarity 95.2%; Pred. No. 1.7e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
DB 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 9  
US-09-919-901-11  
; Sequence 11, Application US/09919901  
; Publication No. US20030082518A1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263,933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: :  
US-09-919-901-11

Query Match 84.5%; Score 93; DB 10; Length 1692;  
Best Local Similarity 95.2%; Pred. No. 1.7e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRILSGRPAIVPRR 23  
DB 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 10

```
US-09-919-901-18
; Sequence 18, Application US/09919901
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-18
Query Match      84.5%; Score 93; DB 10; Length 1692;
Best Local Similarity 95.2%; Pred. No. 1.7e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGRPAIVPDR 23
Db 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 11
US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-10-191-966-4
Query Match      84.5%; Score 93; DB 14; Length 1692;
Best Local Similarity 95.2%; Pred. No. 1.7e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGRPAIVPDR 23
Db 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 12
US-10-191-966-11
; Sequence 11, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-10-191-966-11
Query Match      84.5%; Score 93; DB 14; Length 1692;
Best Local Similarity 95.2%; Pred. No. 1.7e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGRPAIVPDR 23
Db 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 13
US-10-191-966-18
; Sequence 18, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-10-191-966-18
Query Match      84.5%; Score 93; DB 14; Length 1692;
Best Local Similarity 95.2%; Pred. No. 1.7e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GSVVIVGRILSGRPAIVPDR 23
Db 957 GSVVIVGRILSGRPAIVPDR 977

RESULT 14
US-10-085-476-2
; Sequence 2, Application US/10085476
```

/ Publication No. US20020164722A1  
/ GENERAL INFORMATION:  
/ APPLICANT: De Francesco, Raffaele  
/ APPLICANT: Tomei, Licia  
/ APPLICANT: Behrens, Sven-Erik  
/ TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE  
/ TITLE OF INVENTION: RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL  
/ TITLE OF INVENTION: TRANSFERASE ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)  
/ FILE REFERENCE: IT0002PCA  
/ CURRENT APPLICATION NUMBER: US/10/085,476  
/ PRIOR FILING DATE: 2002-02-27  
/ PRIOR APPLICATION NUMBER: 08/952,981  
/ PRIOR FILING DATE: 1998-03-23  
/ PRIOR APPLICATION NUMBER: PCT/IT96/00106  
/ PRIOR FILING DATE: 1996-05-24  
/ PRIOR APPLICATION NUMBER: RM95A000343  
/ PRIOR FILING DATE: 1995-05-25  
/ NUMBER OF SEQ ID NOS: 14  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 2  
/ LENGTH: 2201  
/ TYPE: PRT  
/ ORGANISM: cdna clone pCD (38-9.4)  
US-10-085-476-2

Query Match 84.5%; Score 93; DB 13; Length 2201;  
Best Local Similarity 95.2%; Pred. No. 2.3e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRIILSGRPAIVPDR 23  
|||  
DB 869 GSVVIVGRIILSGRPAIVPDR 889

RESULT 15  
US-09-919-901-2  
/ Sequence 2, Application US/09919901  
/ Publication No. US20030082518A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Potts, Karen E.  
/ APPLICANT: Jackson, Roberta L.  
/ APPLICANT: Patlick, Amy K.  
/ TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
/ TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
/ FILE REFERENCE: 0125-0005A  
/ CURRENT APPLICATION NUMBER: US/09/919,901  
/ CURRENT FILING DATE: 2001-08-02  
/ PRIOR APPLICATION NUMBER: 09/263,933  
/ PRIOR FILING DATE: 1999-02-08  
/ PRIOR APPLICATION NUMBER: 09/129,611  
/ PRIOR FILING DATE: 1998-08-05  
/ NUMBER OF SEQ ID NOS: 33  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 2  
/ LENGTH: 2307  
/ TYPE: PRT  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: :  
US-09-919-901-2

Query Match 84.5%; Score 93; DB 10; Length 2307;  
Best Local Similarity 95.2%; Pred. No. 2.4e-05;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GSVVIVGRIILSGRPAIVPDR 23  
|||  
DB 1049 GSVVIVGRIILSGRPAIVPDR 1069

Search completed: July 6, 2004, 09:40:50  
Job time : 301 secs

